

November 2005

Published\_Applications Nucleic Acid and Published\_Applications Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New). Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

8

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 19:01:43 ; Search time 6169 Seconds  
(without alignments)  
17807.760 Million cell updates/sec

Title: US-10-723-807-1\_COPY\_49\_2396

Perfect score: 2348

Sequence: 1 cctcctctctcttgagcg.....cgacgcgttcgagtcacatg 2348

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_est3.\*

4: gb\_hic.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_est7.\*

9: gb\_gss1.\*

10: gb\_gss2.\*

11: gb\_gss3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	303	12.9	878	6	CF818626 EST696008
2	266	11.3	885	6	CF822062 EST699444
3	239.6	10.2	947	6	CF826270 EST703652
4	204.2	8.7	630	8	DR625376 EST101550
5	154.2	6.6	527	10	CF480293 tigr-gss-
6	146.2	6.2	826	7	CO546284 LyEST239
7	146	6.2	784	9	BZ256150 CH230-428
8	143.8	6.1	620	10	CE793096 tigr-gss-
9	137.6	5.9	590	11	CNS03M2Z
10	137.2	5.8	789	9	BZ115869 CH230-255
11	131	5.6	805	11	CNS04RW2
12	128.6	5.5	774	10	CNS03APC
13	127.2	5.4	572	11	CNS043SN
14	127.2	5.4	723	9	AZ627893
15	125.2	5.3	729	7	CV528119 CS GIL 16
16	123.2	5.2	762	10	AG498356 Mus muscu
17	122.4	5.2	646	7	CO543192
18	120	5.1	633	7	CV070818 CS gil_34
19	118.6	5.1	581	6	CF569196 EST057 Su
20	118.2	5.0	626	9	BZ243402 CH230-419
21	117	5.0	720	11	CR200716 Reverse s
22	117	5.0	812	11	CR072822 Reverse s

#### ALIGNMENTS

##### RESULT 1

##### CF818626

##### LOCUS

##### DEFINITION

CF818626 EST696008 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBB578 3' end, mRNA sequence.

##### ACCESSION

CF818626

##### VERSION

CF818626.1

##### KEYWORDS

EST.

##### SOURCE

Coccidioides posadasii

##### ORGANISM

Coccidioides posadasii

##### REFERENCE

Gardner, M.J. and Cole, G.T.

Analysis of gene expression in Coccidioides posadasii mycelia and

spherules via expressed sequence tags

Unpublished (2003)

Other ESTs: EST696009

Contact: Gardner MJ

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301 838 3519

Fax: 301 838 0208

Email: gardner@tigr.org

Location/Qualifiers

1. 878

/organism="Coccidioides posadasii"

/mol\_type="mRNA"

/strain="C735"

/db\_xref="taxon:199306"

/clone="CIBB578"

/dev\_stage="saprobic phase (mycelia)"

/lab\_host="E. coli DH10B, T1 phage resistant"

/clone\_lib="Coccidioides posadasii saprobic phase cDNA

library, greater than 4kb"

/note="Vector: pExpress 1; Site\_1: Not I; Site\_2: Eco RV;

Coccidioides posadasii saprobic phase cDNA library, size

fractionated cDNA > 4 kb"

##### ORIGIN

Query Match

Best Local Similarity

Matches 523; Conservative

12.9%; Score 303; DB 6; Length 878;

60.2%; Pred. No. 5.2e-67;

Mismatches 340; Indels 6; Gaps 1;





QY 2261 CAGACGAATCCGACGTCGTTGAGTCTGCTTAAATCATCGCTAGCAAGGAGGTGAAGCAG 2320  
Db 595 CAACCAACCCCTACTGTCAGTTCGCTTCGAATGTCGAATCGAATCGAAGACAAAGAAT 654  
QY 2321 GCGCGATCGAGCGCGTTCGAGTCAACTG 2348  
Db 655 GTGACACTTGATGCTTTTGAGCGCGATG 682

RESULT 3  
LOCUS CF826270 947 bp mRNA linear EST 01-APR-2004  
DEFINITION EST703652 Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb Coccidioides posadasii cDNA clone CIDB176 5' end, mRNA

ACCESSION CF826270  
VERSION CF826270.1 GI:45932327  
KEYWORDS Coccidioides posadasii  
SOURCE Coccidioides posadasii  
ORGANISM Coccidioides posadasii  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.

REFERENCE 1 (bases 1 to 947)  
Gardner, M.J. and Cole, G.T.  
Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags  
Unpublished (2003)  
JOURNAL Other ESTs: EST703651  
COMMENT Contact: Gardner MJ  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@tigr.org  
Seq primer: M13 Reverse.

FEATURES  
source  
1..947  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone="CIDB176"  
/dev\_stage="saprobic phase (mycelia)"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_lib="Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb"  
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA 2 to 4 kb"

ORIGIN  
Query Match 10.2%; Score 239.6; DB 6; Length 947;  
Best Local Similarity 65.5%; Pred. No. 1.5e-50;  
Matches 350; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
QY 1815 CGGTGCTGGCGTCCACCAACACCAAGCAGTATCGGCATTCGTTTGGCATTTAT 1874  
Db 3 CGGTTCTGTGGGCAACAAATACAAAAACAGTATCGGCATTCCTTTTGGGATATC 62  
QY 1875 ACGTCCAGATGTGGAGTCACTCAAGTCGCGCCCTGATGCCGATCAGGAAGGTGATA 1934  
Db 63 ACATTCGATTTTGGAAATCTGGCTCAATTGACTCCCTTGACATGACCAACCGGATA 122  
QY 1935 TTCAATATGCCAATCGAGCCCGTGGCAATACCATCGCTTTTGTTCGGAATGACC 1994  
Db 123 TCCAGTACGCTGCTGGAGTCCCAAGGAAATGTTATCGCTATGTCGCAATAACAAT 182  
QY 1995 TTTATCTGGGATATGTTACCGTTCTACGATTAATGATGTTGGCCCCGACATGT 2054  
Db 183 TATACCTCTGGAAGACGGAGGTCACCTCAATACGGAAGATGGAGGCCAAATACTT 242  
2055 TCACGGCGTGGCGGACTGGATCTATGAAGAGGAGATCCTCGCGGATCGTACGCGTTGT 2114

Db 243 TCAACGGTGTTCAGACTGGGTATATGAAGAGGAAATTTTCGGTACTCCGTTTGGTTGT 302  
QY 2115 GGTTCGCCAGATGGTGAATATCTGGCTTACTTGGCTTCAATGAGACTGGGGTTCCGA 2174  
Db 303 GGTTCACCCGACGGCGAATATGTCGTTATAGGAGACCGAGAACTGGCGTCCCAA 362  
QY 2175 CCTACACCGTTTCAGTATTATATGATGAATACCAAGAGATCGTCCGGCGTATCCATGGAGC 2234  
Db 363 CATACACTATCCCATCTACATGATATCAGAGTTTGGCCCTCCGTATCCAGAGAAT 422  
QY 2235 TGAAGATGAATATCCCAAGGTGTCGACAGCATCCGACCGTGCAGTTCGCTTGA 2294  
Db 423 TGGAGCTTCGTTATCTCTAAAGTGTGCAACCAACCTACTGTCAGTTCGCCCTTCTGA 482  
QY 2295 ACATCGCTAGCAAGGAGGTGAAGCAGCGCGCATCGACGCGTTCGAGTCAACTG 2348  
Db 483 ATGTCGAATCCGAGAACAAAGATGTGACACTTGATGCTTTTGAGGGGATG 536

RESULT 4  
LOCUS DR625376 630 bp mRNA linear EST 11-JUL-2005  
DEFINITION DR625376 FvI Gibberella moniliformis cDNA clone FVIBG96, mRNA sequence.

ACCESSION DR625376  
VERSION DR625376.1 GI:70700088  
KEYWORDS Gibberella moniliformis  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE 1 (bases 1 to 630)  
Utterback, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D., Kendra, D.F., Town, C.D. and Whitelaw, C.A.  
Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster  
Unpublished (2005)  
Contact: Brown, D.W.  
USDA/ARS/NCAR

USDA  
1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncar.usda.gov  
TIGR sequence name: FVIBG96TH  
Seq primer: AAT TAA CCC TCA CTA AAG GG.  
Location/Qualifiers  
1..630  
/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="m3125"  
/db\_xref="taxon:117187"  
/clone="FVIBG96"  
/tissue\_type="mycelia"  
/clone\_lib="FvI"  
/note="Vector: pBlueScript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FvI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBlueScript II SK(+) XR vector (CDNA Synthesis Kit; Stratagene)."

FEATURES  
source  
ORIGIN



```
FEATURES
  source
    Location/Qualifiers
      1. 826
      /organism="Petromyzon marinus"
      /mol_type="mRNA"
      /db_xref="taxon:7757"
      /cell_type="lymphocyte"
      /dev_stage="unstimulated larvae"
      /clone_lib="Sea lamprey LyEST"
      /note="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs
from unstimulated larvae. All are from arrayed colonies
from a directionally cloned cDNA library in Lambda ZAP
Express (Stratagene). All are single pass 5' sequences."
ORIGIN
  Query Match      6.2%; Score 146.2; DB 7; Length 826;
  Best Local Similarity 63.0%; Pred. No. 2.4e-26;
  Matches 226; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 449 GTCCGTTGGCGCTGTTGTTGGTGTCTTAGTGGCGGTGGTCTGGCAGTGGTCTGGT 508
DB 472 GCGCGTGAAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 413
QY 509 GCCAAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 568
DB 412 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 353
QY 569 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 628
DB 352 GCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 293
QY 629 GGTGCCAAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 688
DB 292 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 233
QY 689 GTTCTGTGTTGGAGAGCTTCTCAAGTGTCTACTGAGAGTCTTCTAGTGGTGGT 748
DB 232 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 173
QY 749 GGTTCGTGATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 807
DB 172 GGTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 114
RESULT 7
BZ256150/c
LOCUS
DEFINITION
  CH230-428N11.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
  CH230-428N11, genomic survey sequence.
ACCESSION
  BZ256150
VERSION
  BZ256150.1 GI:23962099
KEYWORDS
  GSS.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Rattus.
  1 (bases 1 to 784)
REFERENCE
  AUTHORS
    Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K.,
    Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
    Riggs,F., de Jong,P. and Fraser,C.M.
    Rat BAC End Sequences from Library CHORI-230 MboI segment
    Unpublished (1999)
    Other_GSSs: CH230-428N11.TV
    Contact: Shaying Zhao
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0200
    Fax: 301 838 0208
    Email: szhao@tigr.org
    Clones are derived from the rat BAC library CHORI-230
    (http://www.chori.org/bacpac/rat230.htm). For BAC library
    availability, please contact Pieter de Jong (pdejong@mail.cho.org).
```

```
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/tcdb/bac_ends/rat/bac_end_intro.html
Plate: 428 row: N column: 11
Seq primer: Sp6
Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1. 784
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /strain="BN/SsNHsd/MCW"
      /db_xref="taxon:10116"
      /clone="CH230-428N11"
      /sex="Female"
      /cell_type="Brain"
      /clone_lib="CHORI-230 Segment 2"
      /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
      CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
      Pieter de Jong"
ORIGIN
  Query Match      6.2%; Score 146; DB 9; Length 784;
  Best Local Similarity 58.5%; Pred. No. 2.7e-26;
  Matches 254; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 379 TGGCGGACTTGACGGTGGCAAGGGCGGTCTCGGCCCTTCGTCTGTGGCAGCCAGTAGA 438
DB 703 TGGTATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644
QY 439 CGGTCTCCCTGTGTTGGGCCCTGTTGTTGGTGGTGGTCTTAGGTGGCGGTGGTGGCAG 498
DB 643 TGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 584
QY 499 TGGTCTGTCGCAAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 558
DB 583 TGGTATGATGATGGTATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 524
QY 559 AGACCGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGGTGGTCTTCTAGTGGCGGTGGTGG 618
DB 523 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 464
QY 619 CAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 678
DB 463 TGATGATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 404
QY 679 AGTGACCGTGGTCTGCTGCTGGAGAGCTGCTCAAGTGGTCTACTGGAGGTCTTCTAGG 738
DB 403 TGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 344
QY 739 TGGTATGCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798
DB 343 TGGTATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 284
QY 799 TGCTGGTGGGCAAT 812
DB 283 TGATGGGGGCAAT 270
RESULT 8
CE793096/c
LOCUS
DEFINITION
  tigr-gss-dog-17000317481896 Dog Library Canis familiaris genomic,
  genomic survey sequence.
ACCESSION
  CE793096
VERSION
  CE793096.1 GI:37133863
KEYWORDS
  GSS.
SOURCE
  Canis familiaris (dog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
  Canis.
  1 (bases 1 to 620)
REFERENCE
  AUTHORS
    Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
```

Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (5641), 1898-1903 (2003)  
14512627  
Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.  
Location/Qualifiers  
1. 620  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BatX1; Libraries were prepared from peripheral blood"

[illegible]

RESULT	9
CNS03M2Z	CNS03M2Z
LOCUS	590 bp DNA linear
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 037C13 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL250244
VERSION	AL250244.1 GI:7971256
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis
ORGANISM	Tetraodon nigroviridis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	1	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
AUTHORS		
TITLE		Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL		Nat. Genet. 25 (2), 235-238 (2000)
PUBMED		10835645
REFERENCE	2	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
AUTHORS		
TITLE		Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL		Genome Res. 10 (7), 939-949 (2000)
PUBMED		10899143
REFERENCE	3	(bases 1 to 590)
AUTHORS		Genoscope.
TITLE		Direct Submission
JOURNAL		Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT		This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .
FEATURES		Location/Qualifiers
source		1..590
		/organism="Tetraodon nigroviridis"
		/mol_type="genomic DNA"
		/db_xref="taxon:99883"
		/clone="037C13"
		/clone_lib="G"
		/note="Genoscope sequence ID : C0BG037AB07LP1 end : T7"

[illegible]

```

Db      535  ||||| ||||| ||||| ||||| |||||
          TGGTCTGTATGTTACTGGT 554

RESULT 10
B2115869/c
LOCUS   B2115869.1 789 bp DNA linear GSS 11-OCT-2002
DEFINITION
CH230-255D7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-255D7, genomic survey sequence.
ACCESSION
B2115869
VERSION  B2115869.1 GI:23756816
KEYWORDS
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM
          Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
          1 (bases 1 to 789)
          Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
          Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D.,
          Riggs, F., de Jong, P. and Fraser, C.M.
          Rat BAC End Sequences from Library CHORI-230 MboI segment
          Unpublished (1999)
          Other GSSs: CH230-255D7.TJ
          Contact: Shaving Zhao
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: szhao@tigr.org
          Clones are derived from the rat BAC library CHORI-230
          (http://www.chori.org/bacpac/rat230.htm). For BAC library
          availability, please contact Pieter de Jong (pjejong@mail.cho.org).
          Clones may be purchased from BACPAC Resources
          (http://www.chori.org/bacpac/or ering information.htm). BAC end
          page: http://www.tigr.org/tcdb/bac_ends/rat/bac_end_intro.html
          Plate: 255 row: D column: 7
          Seq primer: 77
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..789
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-255D7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"

ORIGIN
Query Match      5.8%; Score 137.2; DB 9; Length 789;
Best Local Similarity 50.8%; Pred. No. 5.2e-24;
Matches 328; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

QY 174  CTATTGTTGACGGATTGACTGGCGGCAATAAGGGTGGCGAGAGCCTGGAAGCAAGGTTA 233
Db      757  CTGTTGATGTTGCTGATGTTGGTGGTGATGATGCTGATGCTGATGTTGGTGGTGGTCTC 698
QY 234  CTCCTCGTGAAGACCTACCGGAGCGCCCTGATGGCAAGGGCAATGATGGCCCGAGC 293
Db      697  ATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 638
QY 294  GTGATCTTACCGACCTCCGGTCAAGGGGTTTTCACACCTTCGATCTCCCTACTC 353
Db      637  GTGATGATGCTGATGATGTTGCTGATGATGTTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 578
QY 354  CAGAGCTTCTCCGTCGAAGTTCTCTGGCGGACTTCACGGTGGCAAGGGCGGTCTCGGCC 413

```

```

577  GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 518
414  TTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 473
517  CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 458
474  TTCATAGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 533
457  ATGATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 398
534  CCGTTGGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 593
397  GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 338
594  GTGTCCTAGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 653
337  GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 278
654  GTACCCCTAAGCGCGGTGACCGTCCAGTGGAGCGGTTCCTGTCGTTGGAGAGCTTGGTGGTGGT 713
277  GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 218
714  AAGTGTGCTACTGGAGGCTCTCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 773
217  ATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 158
774  ATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTAA 819
157  ATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTAA 112

CNS04RW2      805 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
005C20 of library H from tetraodon nigroviridis, genomic survey
sequence.
AL304427
AL304427.1 GI:8192034
GSS; genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Tetraodon nigroviridis
Tetraodon nigroviridis
Tetraodon nigroviridis
Tetraodon
1 Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
  Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, F.,
  Saurin, W. and Weissenbach, J.
  Estimate of human gene number provided by genome-wide analysis
  using Tetraodon nigroviridis DNA sequence
  Nat. Genet. 25 (2), 235-238 (2000)
2
Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
  Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, F.,
  Saurin, W., Bernot, A. and Weissenbach, J.
  Characterization and repeat analysis of the compact genome of the
  freshwater pufferfish Tetraodon nigroviridis
  Genome Res. 10 (7), 939-949 (2000)
3 (bases 1 to 805)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers

```

**JOURNAL** Submitted (12-APR-2000) Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

**COMMENT** This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

**FEATURES** Location/Qualifiers  
**source**  
1..774  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="252115"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0AG252AE08SP1 end : PUC-Ori"

**ORIGIN**

Query Match	5.5%;	Score 128.6;	DB 10;	Length 774;
Best Local Similarity	58.0%;	Pred. No. 8.8e-22;	Mismatches 175;	Gaps 1
Matches 244;	Conservative	1;		

```

QY      388   TGAACGGTGCAAGGGCGGTCTCGTCGTCTGGTCCGCCAGCCAGTAGACGGTCTCCCC 447
Db      449   TGATGGTGTGATGGTGGTGATGATGATGATGATGATGATGATGATGATGATGATG 390

QY      448   TGTCTGTGGGCCTGTTTGGTGGTCTTCTAGTGGCGGTGGTGGCAGTGGTGGTGGTGG 507
Db      389   TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 330

QY      508   TGGCAAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 567
Db      329   TGATGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 270

QY      568   CCTCTGTTGTTGGCCCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 626
Db      269   TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 210

QY      627   CTGGTGCCAAAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 686
Db      209   ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 150

QY      687   GTGTTCTCTGTTGGAGAGCTTGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 746
Db      149   GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 90

QY      747   CTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806
Db      89   ATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 30

QY      807   G 807
Db      29   G 29

```

**RESULT 13**  
**CNS043SN/c**  
**LOCUS** CNS043SN 572 bp DNA linear GSS 01-SEP-2000  
**DEFINITION** Tetraodon nigroviridis genome survey sequence p7 end of clone 080C12 of library G from Tetraodon nigroviridis, genomic survey sequence.  
**ACCESSION** AL273200  
**VERSION** AL273200.1 GI:7995450  
**KEYWORDS** GSS; genome survey sequence.  
**SOURCE** Tetraodon nigroviridis  
**ORGANISM** Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontoidea; Tetraodontidae; Tetraodon.

**REFERENCE**  
**AUTHORS** Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W., Weissenbach J.  
**TITLE** Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
**JOURNAL** Nat. Genet. 25 (2), 235-238 (2000)  
**PUBMED** 10835645  
**AUTHORS** Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Fizames C., Bernot A., Wincker P., Brottier P., Quetier F., Saurin W., and Weissenbach J.  
**TITLE** Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
**JOURNAL** Genome Res. 10 (7), 939-949 (2000)  
**PUBMED** 10899143  
**AUTHORS** Saurin W., Bernot A., Wincker P., Brottier P., Quetier F., Fizames C., Bernot A., Wincker P., Brottier P., Quetier F., Saurin W., and Weissenbach J.  
**TITLE** Direct Submission



```
Matches 204; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
Qy 471 GTGTTCTAGTGGCGGTGCTGCGCAGTGTGCTGTCGCAAGGTGGTGGTGGTAGTG 530
    |||||
Db 542 GTGTTTCTAGGTTATGTTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 483
    |||||
Qy 531 GTACCGTGGCGTGGCGGACCCAGTACAGTCCCTGTTGTTGGCGCTGTTGTTG 590
    |||||
Db 482 GTGCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 423
    |||||
Qy 591 GTGCTGTCCTAGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 650
    |||||
Db 422 ATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 363
    |||||
Qy 651 GTGTTACCCCTAACGCCGCGTGCACGGTCCAGTGGACGGTTCCTGTCGTTGGAGAGCTTG 710
    |||||
Db 362 GTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 303
    |||||
Qy 711 CTGAAGTGTCTACTGGAGGCTCTTCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 770
    |||||
Db 302 CTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 243
    |||||
Qy 771 CTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 802
    |||||
Db 242 CTGATGTTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 211
    |||||
```

## RESULT 15

```
CV528119 729 bp mRNA linear EST 07-OCT-2004
LOCUS CS_GIL_16F05_SP6 Blue crab gill, normalized Callinectes sapidus
DEFINITION cDNA clone CS_GIL_16F05 5', similar to gb|AA560227.1| phenoloxidase
activating factor - Callinectes sapidus. Score = 140 bits (354),
Expect = 2e-32, mRNA sequence.
```

## ACCESSION

CV528119

CV528119.1 GI:53912498

EST.

Callinectes sapidus (blue crab)

Callinectes sapidus

Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;

Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

Eubrachyura; Portunoidae; Portunidae; Callinectes.

1 (bases 1 to 729)

Shafer, T.H., Coblenz, F.E. and Towle, D.W.

Expressed sequence tags from normalized cDNA libraries prepared

from gill and hypodermis tissues of the blue crab, Callinectes

sapidus

Unpublished (2004)

Contact: Thomas H. Shafer

Department of Biological Sciences

University of North Carolina Wilmington

601 S. College Rd, Wilmington, NC 28403, USA

Tel: 910-962-7275

Fax: 910-962-4066

Email: shafer@uncw.edu

Plate: 16 row: F column: 05

Seq primer: SP6

High quality sequence stop: 508.

Location/Qualifiers

1..729

/organism="Callinectes sapidus"

/mol\_type="mRNA"

/db\_xref="taxon:6763"

/clone="CS\_GIL\_16F05"

/tissue type="Pooled anterior and posterior gills from

crabs acclimated to salinities of 35 and 5 parts per

thousand"

/dev\_stage="Adult intermolt"

/clone.lib="Blue crab gill, normalized"

/note="Vector: pCMV Sport 6.1; Total RNA samples were

prepared individually from each tissue, checked for

quality, and then pooled for construction and

normalization of a cDNA library by Invitrogen. Plasmids

were isolated and inserts sequenced from their 5'-ends by the Blue Crab Molecular Genetics Laboratory at the University of North Carolina Wilmington. Traces were trimmed, compared (BLASTx) to NCBI non-redundant protein database as of 19 July 2004, and processed for submission to dbEST by trace2dbEST software (Parkinson, Anthony and Blaxter, unpublished software)."

## ORIGIN

```
Query Match 5.3%; Score 125.2; DB 7; Length 729;
Best Local Similarity 62.7%; Pred. No. 6.6e-21;
Matches 212; Conservative 0; Mismatches 123; Indels 3; Gaps 1;
Qy 470 GGTGTTCTAGGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 529
    |||||
Db 338 GGCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 397
    |||||
Qy 530 GGTACCGTTGGCGGTGCTGGCAGCCAGTAGACGGTCTCCCTGTTGTTGGCCCTGTTGTT 589
    |||||
Db 398 GGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 457
    |||||
Qy 590 GGTGGTGTCTTAGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 649
    |||||
Db 458 GCGGCTG---GTGGTGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 514
    |||||
Qy 650 AGTGTACCCCTAAGCGCGGTGGACGGTCCAGTGGACGGTCTCTCTGCTTGGAGAGCTT 709
    |||||
Db 515 GGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 574
    |||||
Qy 710 GCTGMAAGGTGCTACTGGAGGTCTTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 769
    |||||
Db 575 GGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 634
    |||||
Qy 770 GCTGATGCTGCTGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 807
    |||||
Db 635 GGTGGAGCTGGTGGTGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 672
    |||||
```

Search completed: January 9, 2006, 00:22:05

Job time : 6175 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 22:39:17 ; Search time 1258 Seconds  
(without alignments)

154334.419 Million cell updates/sec

Title: US-10-723-807-1\_COPY\_49\_2396

Perfect score: 2348

Sequence: 1 cctcctctctctctgagc.....cgacgcgttgagcaactg 2348

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2348	100.0	4280	7	US-10-723-807-1
2	374	15.9	2298	6	US-10-369-493-36489
3	305.4	13.0	2326	9	US-10-926-188A-34
4	169.2	7.2	1493	6	US-10-029-386-25133
5	161	6.9	1635	3	US-09-864-761-20241
6	161	6.9	1973	3	US-09-864-761-3471
7	161	6.9	2543	6	US-10-029-386-20536
8	158	6.7	1075	3	US-09-864-761-19241
9	158	6.7	1403	3	US-09-864-761-2513
10	153.8	6.6	1083	6	US-10-029-386-20723
11	148.2	6.3	1972	6	US-10-017-161-1981
12	148.2	6.3	1972	6	US-10-292-798-1629
13	147	6.3	446	3	US-09-864-761-20699
14	145	6.2	3297	9	US-10-450-763-27241
15	141.6	6.0	79977	9	US-10-737-082-58
16	141.6	6.0	79977	9	US-10-765-790-58
17	138.6	5.9	135259	7	US-10-240-425-1585
18	135	5.7	155074	5	US-10-026-188-6
19	134.8	5.7	439892	5	US-10-087-192-454
20	134	5.7	1395	9	US-10-450-763-25328
21	133.4	5.7	666	6	US-10-029-386-25948
22	133.4	5.7	1543	9	US-10-450-763-28010
23	133.4	5.7	38918	6	US-10-017-161-2049

24	133.4	5.7	38918	6	US-10-292-798-1695	Sequence 1695, Ap
25	133.4	5.7	122279	9	US-10-756-149-3596	Sequence 3596, Ap
26	133	5.7	1200	9	US-10-450-763-28632	Sequence 28632, A
27	132.4	5.6	1231	6	US-10-017-161-2047	Sequence 2047, Ap
28	132.4	5.6	1231	6	US-10-292-798-1693	Sequence 1693, Ap
29	132.2	5.6	1794	9	US-10-450-763-13991	Sequence 13991, A
30	131.8	5.6	461	6	US-10-029-386-20388	Sequence 20388, A
31	128.8	5.5	652	6	US-10-029-386-22783	Sequence 22783, A
32	123.6	5.3	830	3	US-09-864-761-19531	Sequence 19531, A
33	123.6	5.3	2109	9	US-10-450-763-17281	Sequence 17281, A
34	122.6	5.2	479	3	US-09-864-761-768	Sequence 768, App
35	121.8	5.2	744802	6	US-10-292-798-1369	Sequence 1369, Ap
36	120.2	5.1	1168	6	US-10-017-161-2179	Sequence 2179, Ap
37	119.8	5.1	486	6	US-10-465-217-13	Sequence 13, Appl
38	119.4	5.1	39344	5	US-10-087-192-1462	Sequence 1462, Ap
39	118.6	5.1	1168	6	US-10-292-798-1825	Sequence 1825, Ap
40	117.2	5.0	574	3	US-09-864-761-228	Sequence 228, App
41	117.2	5.0	669	3	US-09-864-761-17051	Sequence 17051, A
42	117	5.0	597	6	US-10-029-386-20241	Sequence 20241, A
43	117	5.0	1390	6	US-10-017-161-1781	Sequence 1781, Ap
44	117	5.0	1390	6	US-10-292-798-1437	Sequence 1437, Ap
45	117	5.0	2730	9	US-10-450-763-23586	Sequence 23586, A

#### ALIGNMENTS

##### RESULT 1

US-10-723-807-1  
; Sequence 1, Application US/10723807  
; Publication No. US20040171104A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Blinkovsky  
; APPLICANT: Kimberly Brown  
; APPLICANT: Michael W. Rey  
; APPLICANT: Alan Klotz  
; APPLICANT: Tony Byun  
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl  
; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same  
; FILE REFERENCE: 5254.200-US  
; CURRENT APPLICATION NUMBER: US/10/723,807  
; CURRENT FILING DATE: 2003-11-25  
; PRIOR APPLICATION NUMBER: US/09/079,592  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 08/857,884  
; PRIOR FILING DATE: 1997-05-16  
; PRIOR APPLICATION NUMBER: 60/062,892  
; PRIOR FILING DATE: 1997-10-20  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 4280  
; TYPE: DNA  
; ORGANISM: Aspergillus  
US-10-723-807-1

Query Match	100.0%	Score 2348;	DB 7;	Length 4280;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2348;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CCTCCTCTCTCTTCTGAGCGTTACCGGACTCGTGGCGCTGTCTCCAGTGGCAACGGTAC	60	
Db	49	CCTCCTCTCTCTTCTGAGCGTTACCGGACTCGTGGCGCTGTCTCCAGTGGCAACGGTAC	108	
QY	61	GTATCTTGACGACATGTAAACCGTTGACCTGATGATTAGTGGCCAGCTGGTGGAAAT	120	
Db	109	GTATCTTGACGACATGTAAACCGTTGACCTGATGATTAGTGGCCAGCTGGTGGAAAT	168	
QY	121	CATCGACCGCGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	180	
Db	169	CATCGACCGCGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	228	
QY	181	TGACGATTTGACTGGCGGCAATAAGGGTGGCGAGAGCCTGGAGCAAGGTTACTCTCTCG	240	

Db	229	 TGACGGATTGACTGGCGGCAATAAGGGTGGCGAAGCCTGGAAGCAAGTTACTCTCTCG	288
Qy	241	 TGAAGACCTTACCGGCAGCGCCCTGATGGCAGGGCAATGATGGCCCGCGAGCGTGATCT	300
Db	289	 TGAAGACCTTACCGGCAGCGCCCTGATGGCAAGGGCAATGATGGCCCGCGAGCGTGATCT	348
Qy	301	 TACCGGACGTCCCGGTCAAGGGGTCTTGACAACTCTTCGATCTCTCCCTACTCCAGAGCT	360
Db	349	 TACCGGACGTCCCGGTCAAGGGGTCTTGACAACTCTTCGATCTCTCCCTACTCCAGAGCT	408
Qy	361	 TCCTCCCGTCAAGCTTCTCGCGCACTTGACGCTGGCAAGGGCGGTCTCGGCGTTGCTCG	420
Db	409	 TCCTCCCGTCAAGCTTCTCGCGCACTTGACGCTGGCAAGGGCGGTCTCGGCGTTGCTCG	468
Qy	421	 TCGTGGGAGCCAGTAGACGGTCTCCCTGCTGCTGGCCCTGTTGTTGGTGGTTCCTAGG	480
Db	469	 TCGTGGGAGCCAGTAGACGGTCTCCCTGCTGCTGGCCCTGTTGTTGGTGGTTCCTAGG	528
Qy	481	 TGGCGGTGCTGGCAGTGGTCTGGTGCACAGGCTGGTGGTAGTGGAACCGTTGG	540
Db	529	 TGGCGGTGCTGGCAGTGGTCTGGTGCACAGGCTGGTGGTAGTGGAACCGTTGG	588
Qy	541	 GCGTCTGGCAGCCAGTAGACGGTCTCCCTGTTGTTGGCCCTGTTGTTGGTGGTCTCT	600
Db	589	 GCGTCTGGCAGCCAGTAGACGGTCTCCCTGTTGTTGGCCCTGTTGTTGGTGGTCTCT	648
Qy	601	 AGTGGCGGTGGTCTGGCAGTGGTCTGGTGCACAGGCTGGTGGTAGTGGAACCGTTGG	660
Db	649	 AGTGGCGGTGGTCTGGCAGTGGTCTGGTGCACAGGCTGGTGGTAGTGGAACCGTTGG	708
Qy	661	 TAAGCGCGGTGACGGTCCAGTGGACGGTGTCTGCTGTTGGAGAGCTTGCTGGAAGTGC	720
Db	709	 TAAGCGCGGTGACGGTCCAGTGGACGGTGTCTGCTGTTGGAGAGCTTGCTGGAAGTGC	768
Qy	721	 TACTGGAGGTCTTCTAGGTGGTGTGATGCTGGTCTTCTGCTGATGCTGCTGCTGATGCTGG	780
Db	769	 TACTGGAGGTCTTCTAGGTGGTGTGATGCTGGTCTTCTGCTGATGCTGCTGCTGATGCTGG	828
Qy	781	 TGCTGATGCTGGTGGTGGTGGCAATAGTCTAACAGGGCTTTTACGGCATCAAT	840
Db	829	 TGCTGATGCTGGTGGTGGTGGCAATAGTCTAACAGGGCTTTTACGGCATCAAT	888
Qy	841	 GTGAGGTTATCCAACTCCATCTCTGGTGGCCATTCTGTAATAGCAACAAGAGGGGTGG	900
Db	889	 GTGAGGTTATCCAACTCCATCTCTGGTGGCCATTCTGTAATAGCAACAAGAGGGGTGG	948
Qy	901	 TACTTGGTCCGATGTCAITGCTCTCCGATTTGAAGCTAGCGATTCCTGTATGTACAATA	960
Db	949	 TACTTGGTCCGATGTCAITGCTCTCCGATTTGAAGCTAGCGATTCCTGTATGTACAATA	1008
Qy	961	 ATTTTAAGCAGCTTGGTTCCATFACTGTTCTTCCATGGTTTTTGGATATTTTTTCACTT	1020
Db	1009	 ATTTTAAGCAGCTTGGTTCCATFACTGTTCTTCCATGGTTTTTGGATATTTTTTCACTT	1068
Qy	1021	 ATTGAACTTGTAGTGTCCAGCTTCTCATGGTTTATGATGATGATGATGATGATGATGATGAT	1080
Db	1069	 ATTGAACTTGTAGTGTCCAGCTTCTCATGGTTTATGATGATGATGATGATGATGATGATGAT	1128
Qy	1081	 TCATCTGAGGTTTTGATGTGCAATGGTCAAGTTTTTCTTAAATATGATGATGATGATGATGAT	1140
Db	1129	 TCATCTGAGGTTTTGATGTGCAATGGTCAAGTTTTTCTTAAATATGATGATGATGATGATGAT	1188
Qy	1141	 GGTTACCCGTAGAACTTTGGCCAAATGAAGCTGTAGCTTGTCCAAAGGCTATCAGC	1200
Db	1189	 GGTTACCCGTAGAACTTTGGCCAAATGAAGCTGTAGCTTGTCCAAAGGCTATCAGC	1248
Qy	1201	 GGTTTTTTTTATGAAATCTTTGGCAGATAGGAAAGTGTGGTGGTGGTGGTGGTGGTGGTGGT	1260
Db	1249	 GGTTTTTTTTATGAAATCTTTGGCAGATAGGAAAGTGTGGTGGTGGTGGTGGTGGTGGTGGT	1308
Qy	1261	 GCAGGAGGTGGAGTGTAGTATAAGACGGATTTCTTGGCGGGGAAAGAAAGAAAGCAAT	1320

1309	Db	GCAGGAGGTGGAGTGA	CTGATTAAGACGCGATTTCTCGGGGAAAAAGAAAAGCAACCAAT	1366
1321	Qy	TTATGGGACTATTTATTTAAACGGGAAGTCTTCAATTTCCGCTTTCCGACGCCATCCCTTTGA	1380	
1369	Db	TTATGGGACTATTTATTTAAACGGGAAGTCTTCAATTTCCGCTTTCCGACGCCATCCCTTTGA	1428	
1381	Qy	TTCCGAGCTGAATCTCGGGGTTTTTTTCCACATGAAGTACGTCAATTTCCACTGATTAACA	1440	
1429	Db	TTCCGAGCTGAATCTCGGGGTTTTTTTCCACCATGAAGTACGTCAATTTCCACTGATTAACA	1488	
1441	Qy	TTATTTGTTACATACACTCCATCATTTAGTCAATTAATTAACACCTCATATTTCACTA	1500	
1489	Db	TTATTTGTTACATACACTCCATCATTTAGTCAATTAATTAACACCTCATATTTCACTA	1548	
1501	Qy	CTCAAAGCTTTCTGCTGCTCTCTGCTCAGTGTGGTCCAGGCCCTGGATGTGCTTCGGAAC	1560	
1549	Db	CTCAAAGCTTTCTGCTGCTCTCTGCTCAGTGTGGTCCAGGCCCTGGATGTGCTTCGGAAC	1608	
1561	Qy	ACACGGCCCAACCGGAGAGGCAAGTGAAGCGTCTCACCTTCAATGAGACCGTAGTCAAGCA	1620	
1609	Db	ACACGGCCCAACCGGAGAGGCAAGTGAAGCGTCTCACCTTCAATGAGACCGTAGTCAAGCA	1668	
1621	Qy	AGCAATTAACCGCGACCTCTCGCTCGGTGCAATGCGTCTCGGGCGCAGAGGATGGATCCTA	1680	
1669	Db	AGCAATTAACCGCGACCTCTCGCTCGGTGCAATGCGTCTCGGGCGCAGAGGATGGATCCTA	1728	
1681	Qy	CGTGTACCGGGCGAAGACGGCAGTCTCACCATCGAGAAATCGTCTCAACCAAGAGTCACG	1740	
1729	Db	CGTGTACCGGGCGAAGACGGCAGTCTCACCATCGAGAAATCGTCTCAACCAAGAGTCACG	1788	
1741	Qy	CACGCTCATCCCTGCGGACAAAGATTTCGACAGGGAAGGAAGGTTTCAATTTACTTGGATCCA	1800	
1789	Db	CACGCTCATCCCTGCGGACAAAGATTTCGACAGGGAAGGAAGGTTTCAATTTACTTGGATCCA	1848	
1801	Qy	TCCGCACTTGTGCTCGGTGCTGTGGGCGTCCACCAACCAACGAGTATCGGCATTCGTT	1860	
1849	Db	TCCGCACTTGTGCTCGGTGCTGTGGGCGTCCACCAACCAACGAGTATCGGCATTCGTT	1908	
1861	Qy	CTTTGCGGATTTATTAACGTCAGGATGTGAGTCACTCAAGTCGGTCCGCTGATGCCCGA	1920	
1909	Db	CTTTGCGGATTTATTAACGTCAGGATGTGAGTCACTCAAGTCGGTCCGCTGATGCCCGA	1968	
1921	Qy	TCAGGAAGGTGATTAATCAATATGCCCAATGGAGCCCGTGGGCAATACCAATCGCTTTGT	1980	
1969	Db	TCAGGAAGGTGATTAATCAATATGCCCAATGGAGCCCGTGGGCAATACCAATCGCTTTGT	2028	
1981	Qy	TCGCGAATAGACTTTATGCTCGGATTAATGTTACGTTACTCGCATTTACTGATGATGG	2040	
2029	Db	TCGCGAATAGACTTTATGCTCGGATTAATGTTACGTTACTCGCATTTACTGATGATGG	2088	
2041	Qy	TGGCCCCGACATGTTTCCACGCGTGCCGAGCTGGATCTATGAAGAGGAGATCTCTCGCGCA	2100	
2089	Db	TGGCCCCGACATGTTTCCACGCGTGCCGAGCTGGATCTATGAAGAGGAGATCTCTCGCGCA	2148	
2101	Qy	TCGCTACCGGTTGTGTTCTCGCCAGATGGTGAATATCTGGCTTACTTTGAGCTTCAATGA	2160	
2149	Db	TCGCTACCGGTTGTGTTCTCGCCAGATGGTGAATATCTGGCTTACTTTGAGCTTCAATGA	2208	
2161	Qy	GACTGGGGTCCGACCTACACCGTTCACTATTAATGATGAATCAAGAGATCGCTCCGGC	2220	
2209	Db	GACTGGGGTCCGACCTACACCGTTCACTATTAATGATGAATCAAGAGATCGCTCCGGC	2268	
2221	Qy	GTATCCATGGGAGCTGAAGATAAGGTATCCCAAGGTGTCCGACACGCAATCCGACCGTGAC	2280	
2269	Db	GTATCCATGGGAGCTGAAGATAAGGTATCCCAAGGTGTCCGACACGCAATCCGACCGTGAC	2328	
2281	Qy	GTTGAGTCTGTTTAACATCGCTAGCAAGAGGTGAACAGGCGCGATTCGACCGGTTCCA	2340	
2329	Db	GTTGAGTCTGTTTAACATCGCTAGCAAGAGGTGAACAGGCGCGATTCGACCGGTTCCA	2388	
2341	Qy	GTCAACTG 2348		
2389	Db	GTCAACTG 2396		

Qy	2162	ACTGGGGTTC	CGACCTTAC	CACCGTT	TCAGTAT	TATATG	ATACCA	GAGATCG	CTCCGGC	2222
Db	655	ACTGGCGTAC	CCCACTT	CACCGT	CCAGTACT	TTTATG	GACAA	CCAGGAA	TTGGCCTCCA	714
Qy	2222	TATCCATGG	GAGCTGA	AGATAAG	GTATCC	CAAGGT	TCG	CAGAC	GAATCCG	2281
Db	715	TATCCCGC	CGAGTTC	GCACAT	CAATAC	CCCAAG	GTTC	GAAACA	NAATCCG	774
Qy	2282	TTGAGTCTG	CTTAA	CATCGT	AGCAAG	GGTGA	AGCG	CGCG	CGATCG	2341
Db	775	TTAAATAT	CTCTCAG	CTCAG	CGCAACA	CACCGT	TATCG	ACCATTC	CAATCG	834
Qy	2342	TCAA	2345							
Db	835	CCGA	838							
<p>RESULT 3</p> <p>US-10-926-188A-34</p> <p>; Sequence 34, Application US/10926188A</p> <p>; Publication No. US20050158298A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Atheris Laboratories</p> <p>; TITLE OF INVENTION: Novel Fungal Proteins and Nucleic Acids Encoding Same</p> <p>; FILE REFERENCE: 26511-501</p> <p>; CURRENT APPLICATION NUMBER: US/10/926,188A</p> <p>; CURRENT FILING DATE: 2004-08-25</p> <p>; PRIOR APPLICATION NUMBER: US 60/498,318</p> <p>; PRIOR FILING DATE: 2003-08-25</p> <p>; NUMBER OF SEQ ID NOS: 87</p> <p>; SOFTWARE: Patent In Ver. 2.1</p> <p>; SEQ ID NO 34</p> <p>; LENGTH: 2326</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Trichophyton rubrum</p> <p>; FEATURE:</p> <p>; NAME/KEY: variation</p> <p>; LOCATION: (1835)</p> <p>; OTHER INFORMATION: wherein n can be any nucleotide</p> <p>US-10-926-188A-34</p>										
<p>Query Match 13.0%; Score 305.4; DB 9; Length 2326;</p> <p>Best Local Similarity 61.2%; Pred. No. 5.8e-77;</p> <p>Matches 511; Conservative 1; Mismatches 317; Indels 6; Gaps 1;</p>										
Qy	1511	CTGCTGCTCT	CGTCA	GTGTG	GTGTC	CCAGG	CCCTCG	ATGTG	CTCGG	1570
Db	16	CTACTTATG	TGTGG	CGGCAT	CGCCAA	GCCATCG	TCTCT	CTCGT	GAGCCCGT	75
Qy	1571	ACGGAGAG	GCAGT	TAAGCG	TCTCAC	CTTCAA	TGAG	ACCGT	TAGTCAA	1630
Db	76	ACTGCTGG	CGGCA	CAAGCT	GTGTG	ACCTCA	AGAG	TGTGTCC	CTAGGCT	135
Qy	1631	CCGACCTCT	CGCTCG	GTGCA	ATGCTCT	CGGCG	CGCAG	AGATG	ATCCTAC	1690
Db	136	CCAAGGTC	GACGTCC	CTTGCT	TGATTA	ACAGT	GAAGA	ATGTC	CGGTAT	195
Qy	1691	CGGGAAG	CGCAGT	CTCAC	CATCG	AGAC	CAATCGT	CAAC	ACAGT	1750
Db	196	TCCGAC	ATGGAG	CATTTG	ATCTCT	CGAA	CAATCGT	CA	CGAAC	255
Qy	1751	CTTCGG	CAAGAT	TCCG	ACAGG	GAAG	ACGCTT	CAATTA	CTTGAT	1810
Db	256	GCCGACA	CAAGGT	TACCA	CCAGG	TTACTATG	-----	ACTACT	TGTTCA	309
Qy	1811	TCGTCGG	TGTC	GGCG	CGTCC	AAAC	CAAC	CAAG	CAGTAT	1870
Db	310	TCTGTG	TCTTAT	GGGCA	ACCA	TTATAC	CAAC	AGAGT	ATCCG	369
Qy	1871	TATTAC	GTCC	AGGAT	TGG	GTGCT	CAATG	TC	CGTGC	1930
Db	370	TACTTCA	TCTTAG	ACATCA	AAAGG	ATCG	TTGAC	CCCTCT	TAG	429

1931	Qy	GATATTCAATATGCGCCAAATCGAGCCCGTGGGCAATACCAATCGTGTTTTGTTCGGAGAAT	1990
430	Db	GACATCCAGTATGCTCAATGGAGCCCATGAACAACTCTATCGCCTATGTCCGTGGRAAC	489
1991	Qy	GACCTTATCTCTGGGATATGTTACCGTTACTTCGCATTACTCATGATGATGTGGCCCGAC	2050
490	Db	GACCTGTATATCTGGAACAATGGCAAGACCAAGCGTATTACCGAAATGGCGGCCGGAT	549
2051	Qy	ATGTTCCACCGCGTCCGGAGCTGGAATCTATCAAGAGGAGATCTCTCGCGCATTCGTACGCG	2110
550	Db	ATCTTCAATGGTGTCCCTGACTGGGTATACGAGGAAGAAATCTTCGGGGACCGGTTGCCT	609
2111	Qy	TTGTGGTTCTCGCCAGATGTGTAATATCTCGCTTACTTTGAGCTTCAATAGAGACTGGGGTT	2170
610	Db	CTTTGGTTCTCACCCTGACGGTGAATACCTTTGCGTACCTCCGCTTTTACGAGACTGGAGTC	669
2171	Qy	CCGACCTTACACCGTTCAAGTATTATGGATTAACCAAGAGATCGCTCCGGGGTATCCATGG	2230
670	Db	CCGACCTTACACTATTTCCGTACTACAGAAACAAGCAAAAGATTTGCCCTCTGCTTACCCAAGG	729
2231	Qy	GAGCTGAAGATTAAGGTATCCCAAGTGTCTGCAGACGAATCCGACCGTGAAGTTCAGTCTCG	2290
730	Db	GAGCTGAGATCCGTTTACCTTAAAGTCTCTGCGNAGAACCCACCGTGCAGTTCACACTG	789
2291	Qy	CTTTAAACATCGCTAGCAAGGAGGTGAAGACGGCCGATTCGACGCGTTCGAGTCAA	2345
790	Db	TTAAACATTCGCTTCATCCAGGAGACAACATATCCAGTTACTGTGTTCCCGGAAA	844

## RESULT 4

```

US-10-029-386-25133/c
; Sequence 25133, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25133
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001751.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-25133

```

## RESULT 5

US-09-864-761-20241/C  
 ; Sequence 20241, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 ; FILE REFERENCE: Aemica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 20241  
;; LENGTH: 1635  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC006547.9  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12  
US-09-864-761-20241  
  
Query Match 6.9%; Score 161; DB 3; Length 1635;  
Best Local Similarity 50.6%; Pred. No. 3.6e-35;  
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;  
  
Qy 38 GCTGCTCAGCTGGCAACGGTAGCTATCCTGAACGACAAATGAACCCCTGACTGATGA 97  
Db 1300 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1241  
  
Qy 98 TTATGAGCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 157  
Db 1240 AGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1181  
  
Qy 158 CTTACCAAGGCTCTCCCTATTGTTGACGGATTGACTGGCGGCAATPAAGGGTGGCGAGAAG 217  
Db 1180 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1121  
  
Qy 218 CTTGAAGCAAGTTACTCTCTGTAAGAACCCCTACCGGACGGCCCTGATGGCGAAGGC 277  
Db 1120 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1061  
  
Qy 278 AATGATGCCCGACGGTGAATCTTACCGAGTCCCGGTCAAGGGGCTTTGACAAACCT 337  
Db 1060 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1001  
  
Qy 338 TTGATCTCCCTACTCCAGAGCTTCTCCCGTCAAGCTTCTCTGGCGGACTTGACCGTGGC 397  
Db 1000 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 941  
  
Qy 398 AAGGCGGTCTCGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 457  
Db 940 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881  
  
Qy 458 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 517  
Db 880 GGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 821  
  
Qy 518 GGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 577  
Db 820 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 761  
  
Qy 578 GGGCCCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 637  
Db 760 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 701

Qy 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 697  
Db 700 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 641  
  
Qy 698 GTTGGAGAGCTTGGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 757  
Db 640 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 581  
  
Qy 758 GATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 806  
Db 580 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 532  
  
RESULT 6  
US-09-864-761-3471/c  
; Sequence 3471, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 3471  
; LENGTH: 1973  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006547.9  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11

```

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471
```

```

Query Match          6.9%; Score 161; DB 3; Length 1973;
Best Local Similarity 50.6%; Pred. No. 4.1e-35;
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 38 GCTGCTCCAGTCGCAACGGTACGTATCCTGGAACGCAATGTAAGACGCTTGACTGATGA 97
DB 1454 GGTGATGGTGGTAGTGATGGTGGTGGTAGTGATGGTGGTGGTAGTGATGGTGGT 1395

QY 98 TTAGTAGCCCAAGCTGGTGGAAATCATCGACGGGATCTCCCGTCCCTGTCCTCGACTC 157
DB 1394 AGTCATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 1335

QY 158 CCTACCAAGGCTCCCTATTGTTGACGGATTGACTGGCGCAATAGGGTGGCGAGAAG 217
DB 1334 GGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1275

QY 218 CCTGGAAGCAAGGTTACTCCTCGTGAAGACCTACCGGCAGCGCCCTGATGGCAAGGC 277
DB 1274 GGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGT 1215

QY 278 AATGATGCCCCAGCGGTGATCTTACCCGACGTCGCCGCTCAAGGGGCTTTGACAACT 337
DB 1214 GGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 1155

QY 338 TTGATCTCCCTACTCAGAGCTCTCCCGTCAAGCTTCTGGCGCACTTGACGGTGGC 397
DB 1154 GATGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1095

QY 398 AAGGCGCGTCTCGGCCCTTCGTCGTCGGCAGCCAGTAGACGGTCTCCCTGCTGGTGG 457
DB 1094 GGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1035

QY 458 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 517
DB 1034 GGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 975

QY 518 GGTGCTGAGTGGTACCGTTGGGCGTCGTGGCAGCCAGTAGACGCTCTCCCTGTTGTT 577
DB 974 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 915

QY 578 GGGCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
DB 914 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 855

QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 697
DB 854 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 795

QY 698 GTTGGAGAGCTTGCTGAAGTGCTACTGGAGGCTCTTAGTGGTGGTGGTGGTGGTGG 757
DB 794 GGTGGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 735

QY 758 GATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806
DB 734 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 686
```

```

RESULT 7
US-10-029-386-20536/c
; Sequence 20536, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
```

```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20536
; LENGTH: 2543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
US-10-029-386-20536
```

```

Query Match          6.9%; Score 161; DB 6; Length 2543;
Best Local Similarity 50.8%; Pred. No. 4.8e-35;
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 38 GCTGCTCCAGCTGGCAACGGTACGTATCCTGGAACGCAATGTAAGACGCTTGACTGATGA 97
DB 1283 GGTGATGGTGGTAGTGATGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGT 1224

QY 98 TTAGTAGGCCCAAGCTGGTGGAAATCATCGACCGGATCTTCCCGTCCCTGTCCTCGACTC 157
DB 1223 AGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 1164

QY 158 CCTACCAAGGCTCTCCCTATTGTTGACGGATTGACTGGCGCAATAGGGTGGCGAGAAG 217
DB 1163 GGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1104

QY 218 CCTGGAAGCAAGGTTACTCCTCGTGAAGACCTTACCGGCAGCGCCCTGATGGCAAGGC 277
DB 1103 GGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGT 1044

QY 278 AATGATGGCCCCAGCGGTGATCTTACCGGACGTCGCCGCTCAAGGGGCTTTGACAACT 337
DB 1043 GGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 984

QY 338 TTGATCTCCCTACTCCAGAGCTTCTCCCGTCAAGCTTCTTGGCGACCTTGACGGTGGC 397
DB 983 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 924

QY 398 AAGGCGCGTCTCGGCCCTTCGTCGTCGGCAGCCAGTAGACGGTCTCCCTGTCGTTGGG 457
DB 923 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 864

QY 458 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 517
DB 863 GGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGG 804

QY 518 GGTGCTGAGTGGTACCGTTGGGCGTCGTGGCGACCCAGTAGACGGTCTCCCTGTTGTT 577
DB 803 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 744

QY 578 GGGCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
DB 743 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 684

QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 697
DB 683 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 624

QY 698 GTTGGAGAGCTTGCTGAAGGCTTACTGGAGGCTTCTTAGTGGTGGTGGTGGTGGTGG 757
```





; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2513
; LENGTH: 1403
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL078472.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
; US-09-864-761-2513
Query Match 6.7%; Score 158; DB 3; Length 1403;
Best Local Similarity 53.2%; Pred. No. 2.5e-34;
Matches 335; Conservative 0; Mismatches 295; Indels 0; Gaps 0;
Qy 178 TGTTCACGGATTGACTGCGGCAATAAGGGTGGCGAGAACCTCGAAGCAAGGTTACTCC 237
Db 674 TGGTGTGATGGTATGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 733
Qy 238 TCGTGAAGACCCCTACCGGCACGCCCTGATGGCAAGGGCAATGATGCCGCCGACGGTGA 297
Db 734 TGGTGGTGGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 793
Qy 298 TCTTACGGAGCTCCCGTCAAGGGGTCTTGACACACCTTTTCGATCTCCCTACTCCAGA 357
Db 794 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 853
Qy 358 GCTTCCTCCCTCAAGCTTCTTCGCGGACTTGAACGGTGGCGAAGGCGCGTCTCGGCCCTTCG 417
Db 854 TCGTGGTTTTAGTGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 913
Qy 418 TCGTGTGGGAGCCAGTAGAGGCTCTCCCTGTCTGGGGCTGTGTGTGGTGGTGTCT 477

Db 914 TGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 973
Qy 478 AGGTGCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 537
Db 974 TGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1033
Qy 538 TGGGCGTGGTGGCAGCCAGTAGACCGTCTCCCTGTGTGTGGGCCCTGTGTGGTGGTGT 597
Db 1034 TGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 1093
Qy 598 CCTAGGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 657
Db 1094 TGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 1153
Qy 658 CCCTAAGCCCGTGACCGTCCAGTGACCGGTGTCCTGTGCTGGTGGAGAGCTTGTCTGAAG 717
Db 1154 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1213
Qy 718 TGCTACTGGAGGCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 777
Db 1214 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 1273
Qy 778 TGGTGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 807
Db 1274 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 1303
RESULT 10
US-10-029-386-20723/c
; Sequence 20723, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20723
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049849.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9
; OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-05
; OTHER INFORMATION: EST\_HUMAN HIT: AW794072.1, EVALUE 6.10e+00
; US-10-029-386-20723
Query Match 6.6%; Score 153.8; DB 6; Length 1083;
Best Local Similarity 59.9%; Pred. No. 3.5e-33;
Matches 257; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
Qy 379 TGGCGGACTTGACCGTGGCAAGCGGCTCTCGGCCCTTCGTCGTGTCGTGGCAGCCAGTAGA 438
Db 518 TGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 459
Qy 439 CGGTCTCCCTGTCTGGGCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 498
Db 458 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 399
Qy 499 TGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 558





```
Best Local Similarity 55.18; Pred. No. 2.1e-31;
Matches 299; Conservative 0; Mismatches 243; Indels 1; Gaps 1;

Qy 265 TGATGGCAAGGCAATGATGCCCCGACGGTGATCTTACCGGACGTCCTCCGGTCAAGGGG 324
Db 1593 TGGTGGTGATGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1534
Qy 325 TCTTGACAACCCCTTCATCTCCCTACTCCAGAGCTTCCTCCCGTCAAGCTTCCTGGCGG 384
Db 1533 TGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1474
Qy 385 ACTTGACGGTGGAAGGGCGCTCTCGCCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 444
Db 1473 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1414
Qy 445 CCCTGTCCTTGGGCGCTCTTGTGTGGTGTGTCTAGGTGGCGGTGGTCTGGCAGTGGTC 504
Db 1413 CGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1354
Qy 505 TGGTGCCAAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 563
Db 1353 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1294
Qy 564 GTCTCCCTGTTGTTGGGCGCTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 623
Db 1293 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1234
Qy 624 GTGCTGTGGCAAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 683
Db 1233 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1174
Qy 684 ACGGTGTTCTGTCGTTGGAGAGCTTCGTAAGGTGCTACTGAGAGTCTTCTAGGTGGTG 743
Db 1173 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1114
Qy 744 ATGCTGGTTCGTCGATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 803
Db 1113 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1054
Qy 804 GTG 806
Db 1053 GTG 1051

RESULT 13
US-09-864-761-20699
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007249.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
; OTHER INFORMATION: NT HIT: AL161539.2, EVALUOE 3.70e+00
US-09-864-761-20699

Query Match 6.3%; Score 147; DB 3; Length 446;
Best Local Similarity 60.3%; Pred. No. 1.9e-31;
Matches 243; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 405 GTCTGGCCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 464
Db 7 GTCCCTCAACTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 66
Qy 465 TTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 524
Db 67 GTAGTGGTGGCGGTGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTGGTGG 126
Qy 525 GTAGTGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTGGTGG 584
Db 127 GTGATGGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 186
Qy 585 TTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644
Db 187 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 246
Qy 645 CTGGTAGTGGTACCCCTAAAGCGCGGTACAGTCCAGTGGACGGTTCCTGTCGTTGGAG 704
Db 247 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 306
Qy 705 AGCTTGTCTGAAGGTGCTACTGGAGGTCTTCTAGTGGTGGTGGTGGTGGTGGTGGTGG 764
Db 307 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 366
Qy 765 CTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 807
Db 367 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 409
```



```
Qy 705 AGCTTGCTGAAGGTGCTACTGGAGGCTTCTAGGTGGTGATGCTGGTTCGTGCTGATGCTG 764
    |||||
Db 14747 ATGGTGAGGATGGTGATGGTGATGGTGCGATGGTGATGGTGATGGTGATGGTGATGGTG 14688
    |||||
Qy 765 CTGGTGCTGATGCTGGTGCTGATGCTGGTGCTGGTGCTGGTGCTGGTGCTGGTGCTGG 808
    |||||
Db 14687 GTGATGATGATGATGGCGATGATGGTGAGGATGGTGATGGTGAG 14644
    |||||
```

Search completed: January 9, 2006, 03:12:36  
Job time : 1262 secs











```

; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 169495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (70172)..(70171)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (139457)..(157244)
; OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-61

Query Match 3.8%; Score 88.8; DB 7; Length 169495;
Best Local Similarity 53.4%; Pred. No. 3.5e-11;
Matches 186; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 461 GTTGTGTTGGTGTCTAGTGGCGGTGCTGTCAGTGGTGTGTCGCAAGGCGT 520
DB 67719 GTTGTGTTGGTGTCTAGTGGCGGTGCTGTCAGTGGTGTGTCGCAAGGCGT 67660

QY 521 GCTGTAGTGGTACCGTTGGCGGTGTCGTCAGTGGTGTGTCGCAAGGCGT 580
DB 67659 GTTGTGCGTTTCTGTTGGTGGTGTGTTGTTGTTAGTGTGTTGTTGTTG 67600

QY 581 CCTGTTGTTGGTGTCTAGTGGCGGTGTCGTCAGTGGTGTGTCGCAAGGCGT 640
DB 67599 GTAGTGGTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 67540

QY 641 GGTGCTGTGTAGTGGTACCGTACCGTGTGTCAGTGGTGTGTCGCAAGGCGT 700
DB 67539 ACTGTTGTTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 67480

QY 701 GGAGAGCTTGTGAAGGTGCTACTGAGGTCTTCTAGTGGTGTGTCGCAAGGCGT 760
DB 67479 GACGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 67420

QY 761 GCTGCTGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808
DB 67419 GCTGTTGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 67372

RESULT 10
US-11-136-527-7399/c
; Sequence 7399, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 60/574,294
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 303
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-7399

Query Match 3.6%; Score 85.4; DB 7; Length 2479;
Best Local Similarity 54.2%; Pred. No. 3.6e-11;
Matches 173; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 488 GGTGCTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 547
DB 2301 GCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2242

QY 548 GGCAGCCAGTAGACGGTCTCCCTGTTGTTGGGCGCTGTTGTTGGTGGTGGTGGT 607
DB 2241 GGCATTGCTGCTGTTGGCAGCACTGCTGCTCATTGTTGTAAGTCTCTCTGCTGGC 2182

QY 608 GGTGCTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 667
DB 2181 ACTGCTGCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122

QY 668 GGTGAGGTCCAGTGGACGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
DB 2121 GCTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2062
```



Db 2310 CTGGCTTTACGGGAATGCTGTGTAACTGGTGTGACAGTGTGCTTCCAGGGCTTAAGGGTG 2369  
Qy 396 GCAAGGGCGGTCTCGGCCCTTGTGTGTGAGCAGCCAGTAGAGGTTCTCCCTGCGTTG 455  
Db 2370 ACAGAGGTGATGCTGTGTCCCAAGGTGTGTATGGCTCTCCCTGGCAAGATGGCGTCCGTG 2429  
Qy 456 GGCCTGTTGTTGGTGTGTTCTAGTGTGCGGTGTGTGTGCGAGTGTGTCTGTGTCCCAAGG 515  
Db 2430 GTCTGACCGGCCCAATGTTGTTCTCTGTCGCCCTGTGTGTGTCCTGTGTGCAAGGGTGAA 2489  
Qy 516 GTGTGTCTGTGTAG-----TGCTACCGTTGGCGTGTGTGTGCGAGCCAGTACAGGTTCTCC 569  
Db 2490 GTGTGTCCAGCGGCCCTGTGTGTCTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2549  
Qy 570 CTGTTGTTGGGCTGT 629  
Db 2550 AGCCTGT 2609  
Qy 630 GTGCCAAGGGT 689  
Db 2610 GTGTGTAAAGGGAACCTGT 2669  
Qy 690 TTCTGT 749  
Db 2670 CCGGACCGCTGT 2729  
Qy 750 GTTCTGT 800  
Db 2730 CTGCGGCGAGCGCTGT 2780

RESULT 14  
US-11-136-527-1310  
; Sequence 1310, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AMI01086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 1310  
; LENGTH: 599  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-1310

Query Match 3.5%; Score 82.2; DB 7; Length 599;  
Best Local Similarity 54.4%; Pred. No. 1.1e-10;  
Matches 156; Conservative 4; Mismatches 127; Indels 0; Gaps 0;  
Qy 514 GGGT 573  
Db 175 GGGCTGTGGGGCTGGAACCTGGGGCTGGAGCTGGGGTTGAGGTTGGGGAGGTGTGTGTGTGT 234  
Qy 574 TGTGTGGGCTGT 633  
Db 235 TGT 294  
Qy 634 CAAGGGT 693  
Db 295 TGT 354  
Qy 694 TGT 753  
Db 355 TGT 414  
Qy 754 TGT 800

Db 415 TGCTGT 461  
RESULT 15  
US-11-136-527-5406/c  
; Sequence 5406, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AMI01086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 5406  
; LENGTH: 599  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-5406

Query Match 3.5%; Score 82.2; DB 7; Length 599;  
Best Local Similarity 54.4%; Pred. No. 1.1e-10;  
Matches 156; Conservative 4; Mismatches 127; Indels 0; Gaps 0;  
Qy 514 GGGT 573  
Db 425 GGGCTGTGGGGCTGGAACCTGGGGCTGGAGCTGGGGTTGAGGTTGGGGAGGTGTGTGTGTGT 366  
Qy 574 TGTGTGGGCTGT 633  
Db 365 TGT 306  
Qy 634 CAAGGGT 693  
Db 305 TGT 246  
Qy 694 TGT 753  
Db 245 TGT 186  
Qy 754 TGT 800  
Db 185 TGT 139

Search completed: January 9, 2006, 03:17:05  
Job time : 263 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 12:17:23 ; Search time 937 Seconds  
(without alignments)  
16700.852 Million cell updates/sec

Title: US-10-723-807-1\_COPY\_49\_2396

Perfect score: 2348  
Sequence: 1 cctcctctctctctgagcg.....cgacgcttcgagtcgaactg 2348

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 21:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001as:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002as:\*
- 7: geneseqn2002bs:\*
- 8: geneseqn2003as:\*
- 9: geneseqn2003bs:\*
- 10: geneseqn2003cs:\*
- 11: geneseqn2003ds:\*
- 12: geneseqn2004as:\*
- 13: geneseqn2004bs:\*
- 14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2309.8	98.4	5496	2	AAX07327 Aspergill
2	852.8	36.3	2313	2	AAX00013 Aspergill
3	374	15.9	2298	13	AD560815 Bacterial
4	305.4	13.0	2326	14	ADY51818 T. rubrum
5	220	9.4	491	8	ABZ52886 Aspergill
6	169.2	7.2	1493	12	ACH91938 Human gen
7	161	6.9	1635	4	AAI22780 Probe #12
8	161	6.9	1635	4	ABA67865 Human foe
9	161	6.9	1635	4	AAI48082 Human bre
10	161	6.9	1635	4	ABA49946 Human bre
11	161	6.9	1635	4	ABA34921 Probe #13
12	161	6.9	1635	4	AAK42016 Human bon
13	161	6.9	1635	4	AAI16270 Human bra
14	161	6.9	1635	4	ABA41622 Human liv
15	161	6.9	1635	5	ABA8454 Probe #84
16	161	6.9	1635	6	ABS16047 Human gen
17	161	6.9	1973	4	AAI13556 Probe #34
18	161	6.9	1973	4	ABA55261 Human foe
19	161	6.9	1973	4	AAI34918 Probe #36

c	20	161	6.9	1973	4	ABA44805	Aba44805 Human bre
c	21	161	6.9	1973	4	ABA25005	Aba25005 Probe #34
c	22	161	6.9	1973	4	AAK28970	Aak28970 Human bon
c	23	161	6.9	1973	4	AAK03514	Aak03514 Human bra
c	24	161	6.9	1973	4	ABS28585	Abs28585 Human liv
c	25	161	6.9	1973	5	AAI03446	Aai03446 Probe #34
c	26	161	6.9	1973	6	ABS03504	Abs03504 Human gen
c	27	161	6.9	2543	12	ACH87341	Ach87341 Human gen
c	28	158	6.7	1075	4	AAI21781	Aai21781 Probe #11
c	29	158	6.7	1075	4	ABA66852	Aba66852 Human foe
c	30	158	6.7	1075	4	AAI47064	Aai47064 Probe #15
c	31	158	6.7	1075	4	ABA48937	Aba48937 Human bre
c	32	158	6.7	1075	4	ABA33921	Aba33921 Probe #12
c	33	158	6.7	1075	4	AAK41011	Aak41011 Human bon
c	34	158	6.7	1075	4	AAK15288	Aak15288 Human bra
c	35	158	6.7	1075	4	ABS40603	Abs40603 Human liv
c	36	158	6.7	1075	5	AAI07466	Aai07466 Probe #74
c	37	158	6.7	1075	6	ABS14983	Abs14983 Human gen
c	38	158	6.7	1403	4	AAI12593	Aai12593 Probe #25
c	39	158	6.7	1403	4	ABA54290	Aba54290 Human foe
c	40	158	6.7	1403	4	AAI33941	Aai33941 Probe #26
c	41	158	6.7	1403	4	ABA43834	Aba43834 Human bre
c	42	158	6.7	1403	4	ABA24047	Aba24047 Probe #25
c	43	158	6.7	1403	4	AAK28010	Aak28010 Human bon
c	44	158	6.7	1403	4	AAK02574	Aak02574 Human bra
c	45	158	6.7	1403	4	ABS27605	Abs27605 Human liv

ALIGNMENTS

RESULT 1

AAX07327

ID AAX07327 standard; DNA; 5496 BP.

XX

AC AAX07327;

DT 21-MAY-1999 (first entry)

DE Aspergillus oryzae prolyl-dipeptidyl-peptidase dppiv gene.

XX Prolyl-dipeptidyl-peptidase; dppIV gene; protein hydrolyase; food; ss.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT Promoter 1..1835

FT CDS 1836..4234

FT sig\_peptide 1836..1841

FT exon 1836..1841

FT intron 1842..1924

FT sig\_peptide 1925..1967

FT exon 1943..4234

FT mat\_peptide 1968..4231

FT terminator 4235..4771

XX WO9502705-A1.

XX 21-JAN-1999.

XX 06-MAY-1998; 98WO-EP002799.



[illegible]

Db 361 CATTTCGTTCTTCCGATTATTACGTCAGGATGTCGAGTCACTCAAGTCGCGTCCCGCTG 420  
Qy 1913 ATGCCCATCAGGAAGTGATATTCAATATATGCCCAATGGAGCCCCGTTGGGCATATACCATC 1972  
Db 421 ATGCCCATCAGGAAGTGATATTCAATATATGCCCAATGGAGCCCCGTTGGGCATATACCATC 480  
Qy 1973 GCTTTTCGTCGCGAGATGACCTTTATGTCCTGGGATAATGGTACCGTTACTTCGCATTACT 2032  
Db 481 GCITTTTGTTCGCGAGATGACCTTTATGTCCTGGGATAATGGTACCGTTACTTCGCATTACT 540  
Qy 2033 GATGATGGTGGCCCCGACATGTTCCACGGCGTGCAGGACTGGATCTATGAAGAGGAGATC 2092  
Db 541 GATGATGGTGGCCCCGACATGTTCCACGGCGTGCAGGACTGGATCTATGAAGAGGAGATC 600  
Qy 2093 CTGGGCGATCGTACCGGTTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGGAGC 2152  
Db 601 CTGGGCGATCGTACCGGTTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGGAGC 660  
Qy 2153 TTCAATGAGACTGGGTTCCGACCTACACCGTTCCAGTATTATATGGATAACCAAGAGATC 2212  
Db 661 TTCAATGAGACTGGGTTCCGACCTACACCGTTCCAGTATTATATGGATAACCAAGAGATC 720  
Qy 2213 GCTCCGGCGTATCCATCGGAGCTGAAGATAAGGTATCCCAAGGTGTCGAGACGAATCCG 2272  
Db 721 GCTCCGGCGTATCCATCGGAGCTGAAGATAAGGTATCCCAAGGTGTCGAGACGAATCCG 780  
Qy 2273 ACCGTGACGTTGAGTCTGCTTAAATCGCTAGCAAGAGGTGAAGAGCGGCCGATCGAC 2332  
Db 781 ACCGTGACGTTGAGTCTGCTTAAATCGCTAGCAAGAGGTGAAGAGCGGCCGATCGAC 840  
Qy 2333 GCGTTCGAGTCAACTG 2348  
Db 841 GCGTTCGAGTCAACTG 856

RESULT 3

ADS60815

ID ADS60815 standard; cDNA; 2298 BP.

XX AC ADS60815;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polynucleotide #12802.

XX KW Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.

XX OS Bacteria.

XX FN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

XX PA (HINK/) HINKLE G J.

XX PA (SLAT/) SLATER S C.

XX PA (CHEN/) CHEN X.

XX PA (GOLD/) GOLDMAN B S.

XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX DR WPI; 2004-061375/06.

XX f

PT New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 36489; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 2298 BP; 530 A; 611 C; 655 G; 502 T; 0 U; 0 Other;

Qy Query Match 15.9%; Score 374; DB 13; Length 2298;  
Db Best Local Similarity 66.1%; Pred. No. 5e-75;  
Qy Matches 558; Conservative 0; Mismatches 280; Indels 6; Gaps 1;

Qy 1502 TCCAAGGTTCTGCTCCCTGCTCAGTGTGCTCCAGGCCCTCGGATGCTCGGAAACCA 1561  
Db 1 TCCGTCTCTATCCGTTCTCTCTGCTCAGGTGCTCCAGGCCCGCGCTCCATGGAGACCG 60

Qy 1562 CACGCGCCACCGGAGAGGACGTAAGCGTCTCACCTTCAATGAGACCGTAGTCAAGCAA 1621  
Db 61 CGCGAGCCTAGAGCGCGCGGATCGAAGCGCTTGACATTTCAACGAAACTGTGATTGTCG 120

Qy 1622 GCATTTAGCCGAGCTCTCGCTCGTGCATGCTCTCGGGCGCAGAGATGATCTCTAC 1681  
Db 121 GCGTGTGCGCTTCTTCGATCTCGTGTCAATGGATCGGACTGAGAACGCGGGGATTAT 180

Qy 1682 GTGTACGCGCGGAGAGCGGAGTCTCACCATCGAAGACATCGTCAACACGAGTCAAGC 1741  
Db 181 GTGTACGAGGAGGAGGATGAAGCATCAAGTTGAGAGCATTTGTCACCAACCGATCGAG 240

Qy 1742 ACGCTCATCTCCCTCGGACAAAGATTCCGACAGGAGGAAGCGTTCAATTACTGGATCCAT 1801  
Db 241 ACAATCGTTCCAGCGGAGAGATACCTGC-----CGATCATACAGCTATTGGATTAGC 294

Qy 1802 CCGGACTTGTGTCGGTGTGTCGGGCGTCCAAACACACGAGCGATTCGCATTCGTTTC 1861  
Db 295 CCGGACCTGTCCGCGGTGCTGTGGGCCACGAAATTATTAAGCAGTAGTCCGCCACTCGTTC 354

Qy 1862 TTTGCCGATTTATACGTCGAGGATGTGGAGTCACTCAAGTCCGTCGCCCTGATGCCCGAT 1921  
Db 355 TTGCCGATTTACTACATCCAGGAGCGTCGAGACGTTGGAGACGGTCCGCTTTGGAGGAT 414

Qy 1922 CAGGAAGTGATATTCAATATGCCAATGGAGCCCCCGTGGGCAATACCATCGCTTTTGT 1981  
Db 415 ATGTTGGAGATATTCAATATCGGATGGAGCCCAAGCGGCGATTCCATCGCGTTCGTT 474

Qy 1982 CGCGAGAAATGACCTTTATGTCCTGGGATAATGGTATCCGTTACTCGCATCTGATGATGGT 2041  
Db 475 CGCGCAACAACTTATGGACTTTGCTGTGACGGGACTGTGACGGCTATCACCAGACGGT 534





DE	Aspergillus oryzae polynucleotide SEQ ID NO 1999.	DE	Human genome derived single exon probe #25133.
XX		XX	
KW	Aspergillus oryzae; fermentation; fungus; industrial; EST; expressed sequence tag; gene; ss.	KW	Human; probe; ss; gene expression; single exon probe; microarray; alternative splicing event; genomic alteration.
XX		XX	
OS	Aspergillus oryzae.	OS	Homo sapiens.
XX		XX	
PN	WO200279476-A1.	PN	US2003194704-A1.
XX		XX	
PD	10-OCT-2002.	PD	16-OCT-2003.
XX		XX	
XX	22-MAR-2002; 2002WO-IB000890.	XX	03-APR-2002; 2002US-00029386.
PF		PF	
XX		XX	
XX	30-MAR-2001; 2001JP-00098371.	PR	03-APR-2002; 2002US-00029386.
XX		XX	
PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.	PA	(PENN/) PENN S G.
PA	(NARE-) NAT RES INST BREWING.	PA	(RANK/) RANK D R.
PA	(NORQ ) NAT FOOD RES INST MIN AGRIC.	PA	(HANZ/) HANZEL D K.
XX		XX	
PI	Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H; Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;	PI	Penn SG, Rank DR, Hanzel DK;
XX		XX	
XX	WPI; 2003-046817/04.	DR	WPI; 2004-119264/12.
XX		XX	
XX	Detection of expression of specific Aspergillus genes for monitoring the fermentation and growth conditions of the fungus, using DNA probes.	PT	New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for surveying tissues.
PT		PT	
XX		XX	
PS	Claim 1; SEQ ID NO 1999; 48pp + Sequence Listing; Japanese.	XX	Claim 1; SEQ ID NO 25133; 80pp; English.
XX		PS	
XX	The invention relates to a polynucleotide having any of 6006 specific sequences (ABZ5088-ABZ56893), which are expressed by a fungus under specific culture conditions including one or more of eutrophic, oligotrophic, solid, early germination, alkaline, high temperature, low temperature or maltose culture or polynucleotides stringently hybridising to these sequences. The polynucleotides are useful for monitoring the progress of fermentation and the growth conditions of a fungus, especially of Aspergillus oryzae which is widely used in industrial fermentation. Also monitoring for fungal contamination. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	CC	The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 688 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above, methods of selling and/or licensing single exon probes or microarrays to a customer desiring to measure gene expression, a method of providing human gene expression data by subscription, and a computer-readable storage medium which contains a database having a plurality of records (each record including data on the expression of a single exon probe cited above. The probe, methods and apparatus are useful in gene expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docID=20030194704
CC		CC	
CC	Sequence 491 BP; 114 A; 148 C; 118 G; 111 T; 0 U; 0 Other;	CC	Sequence 1493 BP; 517 A; 800 C; 17 G; 159 T; 0 U; 0 Other;
Qy	Query Match 9.4%; Score 220; DB 8; Length 491; Best Local Similarity 100.0%; Pred. No. 4.5e-40; Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	Query Match 7.2%; Score 169.2; DB 12; Length 1493; Best Local Similarity 54.3%; Pred. No. 2.6e-28; Matches 342; Conservative 0; Mismatches 288; Indels 0; Gaps 0;
Db	104 GGCCAGCTGTGGAAATCATCGACGGGATCTCCCGTCCCTGTCCTGGACTCCCTACC 163       272 GGCCAGCTGTGGAAATCATCGACGGGATCTCCCGTCCCTGTCCTGGACTCCCTACC 331 	Db	178 TGTTGACGGATTGACTGGCGCAATAAGGGTGGCGAGAGCCTGGAAGCAAGTTACTCC 237 
Qy	164 AAGGTCCTCCCTATTGTTGACGGATTGACTGGCGGCAATAAGGGTGGCGAGCCTGGA 223 	Qy	
Db	332 AAGGTCCTCCCTATTGTTGACGGATTGACTGGCGGCAATAAGGGTGGCGAGCCTGGA 391 	Db	
Qy	224 AGCAAGGTTACTCTCGTGAAGACCCCTACCGGCGAGCCCTCGATGGCAAGGCAATGAT 283 	Qy	
Db	392 AGCAAGGTTACTCTCGTGAAGACCCCTACCGGCGAGCCCTCGATGGCAAGGCAATGAT 451 	Db	
Qy	284 GGCCCGGACGGTGATCTTACCGGACGTCCTCCGGTCAAGGGG 323 	Qy	
Db	452 GGCCCGGACGGTGATCTTACCGGACGTCCTCCGGTCAAGGGG 491 	Db	
RESULT 6		XX	
ACH91938/c		XX	
ID ACH91938 standard; DNA; 1493 BP.		XX	
XX		XX	
AC ACH91938;		XX	
XX		XX	
DT 29-JUL-2004 (first entry)		XX	





PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.  
XX Claim 25; SEQ ID NO 16768; 654pp; English.  
XX The present invention relates to single exon nucleic acid probes (SENP).  
XX The present sequence is one such probe. The probes are useful for  
XX producing a microarray for predicting, measuring and displaying gene  
XX expression in samples derived from human placenta. The probes are useful  
XX for antenatal diagnosis of human genetic disorders  
XX  
XX Sequence 1635 BP; 538 A; 916 C; 18 G; 163 T; 0 U; 0 Other;  
Query Match 6.9%; Score 161; DB 4; Length 1635;  
Best Local Similarity 50.6%; Pred. No. 2e-26;  
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;  
QY 38 GCTGCTCAGCTGGCAACGGTACGTATCTGAAACCAATGTAGACGCTTGACTGATGA 97  
DB 1300 GGTGATGGTGTAGTATGATGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1241  
QY 98 TTATGAGCCAGCTGGTGGTAAATCATGACCGCATCTTCCCGTCCCTGCTCCGACTC 157  
DB 1240 AGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1181  
QY 158 CCTACCAAGGGTCTCCCTATTGTTGACGGATTGACTGGCGGCAATAGAGGTGGCGAGAAG 217  
DB 1180 GGTGTGATGGTGTAGTATGGTGTGATGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 1121  
QY 218 CTTGGAAGCAAGTTACTCTCTGTAAGACCCCTACCGCAGCGCCCTGTAGTGGCAAGGC 277  
DB 1120 GGTGTAGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1061  
QY 278 AATGATGCCCGACGGTGTATCTTACCGACGCTCCCGTCAAGGGGTCTTGACAACCT 337  
DB 1060 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1001  
QY 338 TTGCGATCTCCCTACTCCAGAGCTTCTCCCGTCAAGCTTCTCGGGGACTTGACGGTGGC 397  
DB 1000 GATGGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 941  
QY 398 AAGGCGGCTCTCGGCTTCTGTCGTGGCAGCCAGTACAGGCTCTCCCTGCTGCTGGG 457  
DB 940 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881  
QY 458 CTTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 517  
DB 880 GGTAGTGAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 821  
QY 518 GGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 577  
DB 820 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 761  
QY 578 GGGCCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 637  
DB 760 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 701  
QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 697  
DB 700 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 641  
QY 698 GTTGAGAGCTTGCTGAAGCTGCTACTGAGAGGTCTTCTAGGTGGTGGTGGTGGTGGT 757  
DB 640 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 581  
QY 758 GATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806  
DB 580 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 532

RESULT 10  
ABA49946/C  
ID ABA49946 standard; DNA; 1635 BP.  
XX  
AC ABA49946;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human breast cell single exon nucleic acid probe #8641.  
XX  
KW Human; microarray; single exon probe; gene expression; breast; disease;  
XX cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157271-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000662.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-0063236P.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-496933/54.  
XX  
XX New spatially-addressable set of single exon nucleic acid probes, useful  
XX for measuring gene expression in sample derived from human breast,  
XX comprises number of single exon nucleic acid probes.  
XX  
XX Claim 4; SEQ ID NO 8641; 327pp + Sequence Listing; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human breast and BT 474 cells. The method involves contacting the  
XX probes with a collection of detectably labelled nucleic acids derived  
XX from mRNA of human breast, and then measuring the label bound to each  
XX probe of the microarray. The probes are useful for verifying the  
XX expression of regions of genomic DNA predicted to encode proteins. They  
XX are useful for gene discovery, and for determining predisposition and/or  
XX assessing the toxicity of chemical agents on cells. The microarray of  
XX this invention presents a far greater diversity of probes for measuring  
XX gene expression, with far less bias than expressed sequence tag  
XX microarrays. The method is suitable for rapid production of functional  
XX information from genomic sequence. The present sequence is a single exon  
XX nucleic acid probe of the invention. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1635 BP; 538 A; 916 C; 18 G; 163 T; 0 U; 0 Other;  
Query Match 6.9%; Score 161; DB 4; Length 1635;  
Best Local Similarity 50.6%; Pred. No. 2e-26;  
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;  
QY 38 GCTGCTCAGCTGGCAACGGTACGTATCTGAAACCAATGTAGACGCTTGACTGATGA 97  
DB 1300 GGTGATGGTGTAGTATGATGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1241  
QY 98 TTATGAGCCAGCTGGTGGTAAATCATGACCGCATCTTCCCGTCCCTGCTCCGACTC 157









```
QY 98 TTAGTAGCCAGCTGGTGGAAATCATGACCGCATCTTCCCGTCCCTGTCCCTGGACTC 157
Db 1240 AGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1181
QY 158 CCTACCAAGGCTCTCCCTATTGTTGACCGAATTGACTGGCGGCAATAAGGGTGGCGAGAAG 217
Db 1180 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1121
QY 218 CCTGGAACAGAGTTACTCTCGTGAAGACCCCTACCGGACGCGCCCTGTATGGCAGGGC 277
Db 1120 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1061
QY 278 AATGATGCCCGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 337
Db 1060 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1001
QY 338 TTGATCTCCCTACTCCAGAGCTTCTCCCTCCGCTCAAGCTTCTCTGGCGGACTTGAAGGTGGC 397
Db 1000 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 941
QY 398 AAGGCGGTCTGGCTTCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 457
Db 940 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881
QY 458 CTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 517
Db 880 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 821
QY 518 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 577
Db 820 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 761
QY 578 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 637
Db 701 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 701
QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 697
Db 700 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 641
QY 698 GTTGAGAGCTTGTGGAAGTCTACTTGAAGTCTTCTAGGTGGTGGTGGTGGTGGTGGTGGT 757
Db 640 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 581
QY 758 GATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806
Db 580 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 532
```

## RESULT 15

AA108454/c  
ID AA108454 standard; DNA; 1635 BP.

XX AC AA108454;

XX DT 09-OCT-2001 (first entry)

XX DE Probe #8445 used to measure gene expression in human breast sample.

XX KW Probe; human; breast disease; cancer; development disorder; as;  
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX OS Homo sapiens.  
XX PN WO200157270-A2.  
XX PD 09-AUG-2001.  
XX PF 29-JAN-2001; 2001WO-US0000661.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WIPI; 2001-476286/51.  
XX DR Novel single exon nucleic acid probe used to measuring gene expression in  
XX PT a human breast.  
XX PS Claim 25; SEQ ID NO 8445; 322pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes.  
XX CC The present sequence is one such probe. The probes are useful for  
XX CC measuring human gene expression in a human breast sample, where the probe  
XX CC hybridises at high stringency to a nucleic acid expressed in the human  
XX CC breast. The probes are useful for predicting, diagnosing, grading,  
XX CC staging, monitoring and prognosing diseases of the human breast,  
XX CC particularly those diseases with polygenic aetiology. The diseases  
XX CC include: breast cancer, disorders of development, inflammatory diseases  
XX CC of the breast, fibrocystic changes, proliferative breast disease and non-  
XX CC carcinoma tumours. Note: The sequence data for this patent did not form  
XX CC part of the printed specification, but was obtained in electronic format  
XX CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 1635 BP; 538 A; 916 C; 18 G; 163 T; 0 U; 0 Other;

Query Match 6.9%; Score 161; DB 5; Length 1635;

Best Local Similarity 50.6%; Pred. No. 2e-26;

Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 38 GCTGCTCCAGCTGGCAACGCTACCTCTGAAACGACAATGTAAAGCCTTGACTGATGA 97  
Db 1300 GGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1241

QY 98 TTAGTAGCCAGCTGGTGGAAATCATGACCGCATCTTCCCGTCCCTGTCCCTGGACTC 157  
Db 1240 AGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1181

QY 158 CCTACCAAGGCTCTCCCTATTGTTGACCGAATTGACTGGCGGCAATAAGGGTGGCGAGAAG 217  
Db 1180 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1121

QY 218 CTGGAAGCAAGGTTACTCTCTGTAAGACCCCTACCGGACGCGCCCTGTATGGCAGGGC 277  
Db 1120 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1061

QY 278 AATGATGCCCGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 337  
Db 1060 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1001

QY 338 TTGATCTCCCTACTCCAGAGCTTCTCCCTCCGCTCAAGCTTCTCTGGCGGACTTGAAGGTGGC 397  
Db 1000 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 941

QY 398 AAGGCGGTCTGGCTTCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 457  
Db 940 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881

QY 458 CTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 517  
Db 880 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 821

QY 518 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 577  
Db 820 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 761

QY 578 GGGCCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 20:01:38 ; Search time 298 Seconds  
(without alignments)

14005.754 Million cell updates/sec

Title: US-10-723-807-1\_COPY\_49\_2396

Perfect score: 2348

Sequence: 1 cctctctctctctgagc.....cgacgcttcgagcaactg 2348

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 88870828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2348	100.0	4280	3	US-09-079-592-1
2	2309.8	98.4	5496	3	US-09-462-284-1
3	161.8	6.9	152132	3	US-09-949-016-13845
4	161.8	6.9	152145	3	US-09-949-016-12371
5	147	6.3	34230	3	US-09-949-016-12052
6	147	6.3	128470	3	US-09-949-016-13765
7	145	6.2	39154	3	US-09-949-016-12384
8	145	6.2	39154	3	US-09-949-016-12801
9	145	6.2	39443	3	US-09-949-016-14326
10	145	6.2	39443	3	US-09-949-016-14327
11	128.2	5.5	522	3	US-09-949-016-103758
12	127	5.4	30656	3	US-09-949-016-14613
13	121.8	5.2	10304	3	US-09-627-465B-1
14	119.8	5.1	486	3	US-09-639-207-13
15	118.4	5.0	767677	3	US-09-949-016-12147
16	118.4	5.0	767677	3	US-09-949-016-17361
17	116.4	5.0	114793	3	US-10-148-806-3
18	114.2	4.9	55195	3	US-09-949-016-15954
19	112.8	4.8	142783	3	US-09-949-016-15127
20	111.6	4.8	57280	3	US-09-949-016-11796
21	111.6	4.8	57280	3	US-09-949-016-12843
22	111.6	4.8	57280	3	US-09-949-016-12844
23	111.6	4.8	57280	3	US-09-949-016-12846
24	111.6	4.8	57280	3	US-09-949-016-13542

#### ALIGNMENTS

##### RESULT 1

US-09-079-592-1  
; Sequence 1, Application US/09079592B  
; Patent No. 6864092

##### GENERAL INFORMATION:

APPLICANT: Alexander Blinkovsky  
APPLICANT: Kimberly Brown  
APPLICANT: Michael W. Rey  
APPLICANT: Alan Klotz  
APPLICANT: Tony Byun

TITLE OF INVENTION: Polypeptides Having Dipeptidyl

TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same

FILE REFERENCE: 5254.200-US

CURRENT APPLICATION NUMBER: US/09/079,592B

CURRENT FILING DATE: 1998-05-15

PRIOR APPLICATION NUMBER: 08/857,884

PRIOR FILING DATE: 1997-05-16

PRIOR APPLICATION NUMBER: 60/062,892

PRIOR FILING DATE: 1997-10-20

NUMBER OF SEQ ID NOS: 13

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 4280

TYPE: DNA

ORGANISM: Aspergillus

US-09-079-592-1

Query Match 100.0%; Score 2348; DB 3; Length 4280;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCTCTCTCTCTCTGAGCGTTACCGGACTCGTCGCGCTGCCAGCTCCAGCTGGCAACGGTAC	60
DB	49	CCTCTCTCTCTCTGAGCGTTACCGGACTCGTCGCGCTGCCAGCTCCAGCTGGCAACGGTAC	108
QY	61	GTATCTCTGAACGACCAATGTAGACGCTTGACTGATGTAGTGGCCAGCTGGTGAAT	120
DB	109	GTATCTCTGAACGACCAATGTAGACGCTTGACTGATGTAGTGGCCAGCTGGTGAAT	168
QY	121	CATGACCGCGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGCTCCCTATTGT	180
DB	169	CATGACCGCGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGCTCCCTATTGT	228
QY	181	TCACGGATTGATGGCGGCAATAAGGGTGGCGAAGCTTGAAGCAAGTTACTCTCTCG	240
DB	229	TCACGGATTGATGGCGGCAATAAGGGTGGCGAAGCTTGAAGCAAGTTACTCTCTCG	288

QY 241 TGAAGACCCCTACCGGACGCGCCCTGATGGCAAGGGCAATGATGGCCCCCGACCGTGATCT 300  
DB 289 TGAAGACCCCTACCGGACGCGCCCTGATGGCAAGGGCAATGATGGCCCCCGACCGTGATCT 348  
QY 301 TACGGAGCTCCCGGTCAAGGGGTCTTGACAAACCCCTTCGATCTCCCTACTCCAGAGCT 360  
DB 349 TACGGAGCTCCCGGTCAAGGGGTCTTGACAAACCCCTTCGATCTCCCTACTCCAGAGCT 408  
QY 361 TCCTCCCGTCAAGCTTCCTCGCGACTTGACGGTGGCAAGGGGGTCTCGGCCCTTCGTG 420  
DB 409 TCCTCCCGTCAAGCTTCCTCGCGACTTGACGGTGGCAAGGGGGTCTCGGCCCTTCGTG 468  
QY 421 TCGTGGCAGCCAGTAGACGGTCTCCCTGTCTGCTGGTGGGCCCTGTGTTGGTGGTTCAGG 480  
DB 469 TCGTGGCAGCCAGTAGACGGTCTCCCTGTCTGCTGGGCCCTGTGTTGGTGGTTCAGG 528  
QY 481 TGGCGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTACCGTTGG 540  
DB 529 TGGCGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTACCGTTGG 588  
QY 541 GCCTCGTGGCAGCCAGTAGACGGTCTCCCTGTGTTGGGCCCTGTGTTGGTGGTTCCT 600  
DB 589 GCCTCGTGGCAGCCAGTAGACGGTCTCCCTGTGTTGGGCCCTGTGTTGGTGGTTCCT 648  
QY 601 AGTGGCGGTGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTACCCC 660  
DB 649 AGTGGCGGTGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTACCCC 708  
QY 661 TAGCGCGGTGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTGGTGC 720  
DB 709 TAGCGCGGTGGTGGTGGCAGTGGTCTGCTGGTGGCAAGGGTGGTGGTGGTGGTGGTGC 768  
QY 721 TACTGGAGGTCTCTAGGTGGTGGTGGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 780  
DB 769 TACTGGAGGTCTCTAGGTGGTGGTGGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 828  
QY 781 TGTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840  
DB 829 TGTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 888  
QY 841 GTGAGGTATCCACATCCATCTCTGGTGGCCATTCGTAATAGCAACAAGGGGTGG 900  
DB 889 GTGAGGTATCCACATCCATCTCTGGTGGCCATTCGTAATAGCAACAAGGGGTGG 948  
QY 901 TACTTGGTCGCGATGTCAATGCTCCTCGGATTTGAAGCTAGCGATTCTGTATGTACAA 960  
DB 949 TACTTGGTCGCGATGTCAATGCTCCTCGGATTTGAAGCTAGCGATTCTGTATGTACAA 1008  
QY 961 ATTTTAAGCAGCTTGGTTCATCTGTTCTTCACTGGTTTTGGGATATTTTTTCACTT 1020  
DB 1009 ATTTTAAGCAGCTTGGTTCATCTGTTCTTCACTGGTTTTGGGATATTTTTTCACTT 1068  
QY 1021 ATTGAATCTTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080  
DB 1069 ATTGAATCTTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1128  
QY 1081 TCATCTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1140  
DB 1129 TCATCTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1188  
QY 1141 GGTATACCGGTAGAGCTTTGCCAATGAAGCTGTAGCTTGTCCAAAGGGTATCAGC 1200  
DB 1189 GGTATACCGGTAGAGCTTTGCCAATGAAGCTGTAGCTTGTCCAAAGGGTATCAGC 1248  
QY 1201 GGTATATATGAGATCTTGGCAGGATAGGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1260  
DB 1249 GGTATATATGAGATCTTGGCAGGATAGGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1308  
QY 1261 GCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1320  
DB 1309 GCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1368  
QY 1321 TTATGGGACTATTTTAAACGGGAAGTCTTCAATTCGGCTTCGCCAGCCATCCCTTGA 1380

RESULT 2

DB 1369 TTATGGGACTATTTTAAACGGGAAGTCTTCAATTCGGCTTCGCCAGCCATCCCTTGA 1428  
QY 1381 TTCGAGCTGAACCTCGGGTTTTTCCACCATGAAGGTAGTCAATTCACATGATTAACA 1440  
DB 1429 TTCGAGCTGAACCTCGGGTTTTTCCACCATGAAGGTAGTCAATTCACATGATTAACA 1488  
QY 1441 TTATTTGTTACATACACTCCATCATTTAGTCAATTAATTAACACCTCATTAATTCAGTA 1500  
DB 1489 TTATTTGTTACATACACTCCATCATTTAGTCAATTAATTAACACCTCATTAATTCAGTA 1548  
QY 1501 CTCGAAGCTTCGCTCTCTGCTGCTGAGTGGTCCAGGGCCTGATGTGCTCGGAAACC 1560  
DB 1549 CTCGAAGCTTCGCTCTCTGCTGCTGAGTGGTCCAGGGCCTGATGTGCTCGGAAACC 1608  
QY 1561 ACACGGCCGCCACCGGAGAGGAGTAAAGCGTCTCACCTTCATTAAGAGACCGTAGTCAAGCA 1620  
DB 1609 ACACGGCCGCCACCGGAGAGGAGTAAAGCGTCTCACCTTCATTAAGAGACCGTAGTCAAGCA 1668  
QY 1621 AGCAATTAAGCCGACCTCTCGCTCGGTGCAATGGCTCTCGGGGCGCAGAGGATGATCCTA 1680  
DB 1669 AGCAATTAAGCCGACCTCTCGCTCGGTGCAATGGCTCTCGGGGCGCAGAGGATGATCCTA 1728  
QY 1681 CGGTACGGCGGCGAAGAGCGCAGTCTCACCATCGAGAACATCGTCAACCAACGAGTCAAG 1740  
DB 1729 CGGTACGGCGGCGAAGAGCGCAGTCTCACCATCGAGAACATCGTCAACCAACGAGTCAAG 1788  
QY 1741 CACGCTCATCCCTCGCGACAAAGATTCGACAGGAAAGGAGCGTTCAATTTACTGGATGCCA 1800  
DB 1789 CACGCTCATCCCTCGCGACAAAGATTCGACAGGAAAGGAGCGTTCAATTTACTGGATGCCA 1848  
QY 1801 TCCGAGCTTGTGCTCGGTGCTGGGGTCCAAACCAACCAAGCAGTATCGGCATTCGTT 1860  
DB 1849 TCCGAGCTTGTGCTCGGTGCTGGGGTCCAAACCAACCAAGCAGTATCGGCATTCGTT 1908  
QY 1861 CTTTGCGGATTTTACGTCCAGGATGTGGAGTCACTCAAGTCCGTGCCCTGATGCCCGA 1920  
DB 1909 CTTTGGCGATTTTACGTCCAGGATGTGGAGTCACTCAAGTCCGTGCCCTGATGCCCGA 1968  
QY 1921 TCAGGAAGGTGATATTTCAATATGCCCAATGGAGCCCCCGTGGGCAATAACCATCGCTTTGT 1980  
DB 1969 TCAGGAAGGTGATATTTCAATATGCCCAATGGAGCCCCCGTGGGCAATAACCATCGCTTTGT 2028  
QY 1981 TCAGGAAGGTGATATTTCAATATGCCCAATGGAGCCCCCGTGGGCAATAACCATCGCTTTGT 2040  
DB 2029 TCAGGAAGGTGATATTTCAATATGCCCAATGGAGCCCCCGTGGGCAATAACCATCGCTTTGT 2088  
QY 2041 TGGCCCCGACATGTTCCACGGCGTGGCGGACTGGATCTATGAAGAGGAGATCCTCGCGCA 2100  
DB 2089 TGGCCCCGACATGTTCCACGGCGTGGCGGACTGGATCTATGAAGAGGAGATCCTCGCGCA 2148  
QY 2101 TCGCTACCGCTTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGAGCTTCAATGA 2160  
DB 2149 TCGCTACCGCTTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGAGCTTCAATGA 2208  
QY 2161 GACTGGGGTTCCGACCTTACACCGTTTCAATTTATATGGATTAACCAAGAGATCGCTCCGGC 2220  
DB 2209 GACTGGGGTTCCGACCTTACACCGTTTCAATTTATATGGATTAACCAAGAGATCGCTCCGGC 2268  
QY 2221 GTATCCATGGAGCTGAAGATTAAGGTATCCCAAGGTGTCCGACAGCAATCCGACCGGTGAC 2280  
DB 2269 GTATCCATGGAGCTGAAGATTAAGGTATCCCAAGGTGTCCGACAGCAATCCGACCGGTGAC 2328  
QY 2281 GTTGAAGTCTCTTAAACATCGCTAGCAAGGAGGTGAAGAGCGCGCGGATCGACGGTTCGA 2340  
DB 2329 GTTGAAGTCTCTTAAACATCGCTAGCAAGGAGGTGAAGAGCGCGCGGATCGACGGTTCGA 2388  
QY 2341 GTCAACTG 2348  
DB 2389 GTCAACTG 2396

US-09-462-284-1  
; Sequence 1, Application US/09462284  
; Patent No. 630988  
; GENERAL INFORMATION:  
; APPLICANT: Nestec S.A.  
; APPLICANT: Monod, Michel  
; APPLICANT: Doumas, Agnes  
; APPLICANT: Affolter, Micheal  
; APPLICANT: Van Den Broek, Peter  
; TITLE OF INVENTION: CLONING OF THE  
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM  
; TITLE OF INVENTION: ASPERGILLUS ORYZAE  
; FILE REFERENCE: 8265-298  
; CURRENT APPLICATION NUMBER: US/09/462,284  
; CURRENT FILING DATE: 2000-01-03  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5496  
; TYPE: DNA  
; ORGANISM: Fungus  
US-09-462-284-1

Query Match 98.4%; Score 2309.8; DB 3; Length 5496;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2344; Conservative 0; Mismatches 2; Indels 3; Gaps 3;  
QY 1 GCTCTCTCTCTCTGAGGGTTACCGGACTCGTCCCGCTGCTCAGCTGGCAAGGTAC 60  
DB 428 GCTCTCTCTCTCTGAGGGTTACCGGACTCGTCCCGCTGCTCAGCTGGCAAGGTAC 487  
QY 61 GATCCTGAAACGACAAATGTAAGACGCTTGACTGATGATTAGTAGGCCAGCTGGTGGAAAT 120  
DB 488 GATCCTGAAACGACAAATGTAAGACGCTTGACTGATGATTAGTAGGCCAGCTGGTGGAAAT 547  
QY 121 CATCAGCCGCGATCTTCCCGTCCCTGCTCCCTGGAATCCCTACCAAGGCTCCCTATTGT 180  
DB 548 CATCAGCCGCGATCTTCCCGTCCCTGCTCCCTGGAATCCCTACCAAGGCTCCCTATTGT 607  
QY 181 TGACGGATTGATGGCGGCMATAAGGGTGGCGAAGCTTGAAGAGCTTACTCCCTCG 240  
DB 608 TGACGGATTGATGGCGGCMATAAGGGTGGCGAAGCTTGAAGAGCTTACTCCCTCG 667  
QY 241 TGAAGACCTACCGGCGCCCTGATGGCAAGGCAATGATGCGCGAGCGGTGATCT 300  
DB 668 TGAAGACCTACCGGCGCCCTGATGGCAAGGCAATGATGCGCGAGCGGTGATCT 727  
QY 301 TACCGAGCTCCCGGCTCAAGGGGGCTTGAACAACCTTTTCGATCTCCCTACTCCAGAGCT 360  
DB 728 TACCGAGCTCCCGGCTCAAGGGGGCTTGAACAACCTTTTCGATCTCCCTACTCCAGAGCT 787  
QY 361 TCTCCCGTCAAGCTTCTCGGAGACTTGAACGGGCTTGAACGGGCTTCCGCTTCGTCG 420  
DB 788 TCTCCCGTCAAGCTTCTCGGAGACTTGAACGGGCTTGAACGGGCTTCCGCTTCGTCG 847  
QY 421 TCGTGGCAGCCAGTAGAGCGTCTCCCTGCTGGGCTGTTGTTGGTGGTCTAGG 480  
DB 848 TCGTGGCAGCCAGTAGAGCGTCTCCCTGCTGGGCTGTTGTTGGTGGTCTAGG 907  
QY 481 TGGCGGTGCTGCTGGCAGTGGTCTGGTCCCAAGGGTGGTGGTGGTGGTGGTGGTGG 540  
DB 908 TGGCGGTGCTGCTGGCAGTGGTCTGGTCCCAAGGGTGGTGGTGGTGGTGGTGGTGG 967  
QY 541 GCGTGGTGGCAGCCAGTAGAGCGTCTCCCTGCTGGGCTGTTGTTGGTGGTGGTGGT 600  
DB 968 GCGTGGTGGCAGCCAGTAGAGCGTCTCCCTGCTGGGCTGTTGTTGGTGGTGGTGGT 1027  
QY 601 AGGTGGCGGTGGTGGTGGCAGTGGTCTGGTCCCAAGGGTGGTGGTGGTGGTGGTGG 660  
DB 1028 AGGTGGCGGTGGTGGTGGCAGTGGTCTGGTCCCAAGGGTGGTGGTGGTGGTGGTGG 1087  
QY 661 TAAAGCCCGTGAAGGCTCCAGTGGAGCGGTTCCTGCTGGAGAGCTTGGTGAAGGTGC 720

DB 1088 TAAGCGCGGTGACGGTCCAGTGGACGGTGTCTGCTGTTGGAGAGCTTGTCTGAAGGTGC 1147  
QY 721 TACTGGAGGTCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780  
DB 1148 TACTGGAGGTCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1207  
QY 781 TGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840  
DB 1208 TGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1267  
QY 841 GTGAGGTTATCCAAACATCCATCCCTTGGTGGGCAATTCGTAATAGCAACAAAGAGGGGTGG 900  
DB 1268 GTGAGGTTATCCAAACATCCATCCCTTGGTGGGCAATTCGTAATAGCAACAAAGAGGGGTGG 1327  
QY 901 TACTTGGTGGGATGTCAATGCTCTCTCGATTTGAAGCTAGCGATTCCTGTATGATCAATA 960  
DB 1328 TACTTGGTGGGATGTCAATGCTCTCTCGATTTGAAGCTAGCGATTCCTGTATGATCAATA 1387  
QY 961 ATTTTAAGCAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
DB 1388 ATTTTAAGCAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1447  
QY 1021 ATTGAATCTTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080  
DB 1448 ATTGAATCTTGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1507  
QY 1081 TCATCTGAGGTTGATGTTGCAATGGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1140  
DB 1508 TCATCTGAGGTTGATGTTGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1567  
QY 1141 GGTATACCGCTAGAGGCTTGGCCACCAATGAAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1200  
DB 1568 GGTATACCGCTAGAGGCTTGGCCACCAATGAAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1627  
QY 1201 GGTATATATGAGAACTTGGGAGGATAGAAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1260  
DB 1628 GGTATATATGAGAACTTGGGAGGATAGAAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1687  
QY 1261 GCAGAGGTGGAGTGAATGAAGCGGATTTTCGCGGGGAAAAGAAAAGAGGAGCAAT 1320  
DB 1688 GCAGAGGTGGAGTGAATGAAGCGGATTTTCGCGGGGAAAAGAAAAGAGGAGCAAT 1747  
QY 1321 TTATGGGACTATTTTAAACGGGAACTCTTCAATTCGGCTTCGGCAGCCATCCCTTGA 1380  
DB 1748 TTATGGGACTATTTTAAACGGGAACTCTTCAATTCGG- TTCCGAGCCATCCCTTGA 1806  
QY 1381 TTCGAGCTGAACCTCGGGGTTTTCACCATGAAGTACCTCAATTCACATGATTAATAACA 1440  
DB 1807 TTCGAGCTGAACCTCGGGGTTTTCACCATGAAGTACCTCAATTCACATGATTAATAACA 1866  
QY 1441 TTATTTGTTACATACACTCCATCATTTGAGTCAATTAATTAACACCTCATTAATTCAGTA 1500  
DB 1867 TTATTTGTTACATACACTCCATCATTTGAGTCAATTAATTAACACCTCATTAATTCAGTA 1926  
QY 1501 TTCCAAAGCTTCTGCTGCTCCTGGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1560  
DB 1927 TTCCAAAGCTTCTGCTGCTCCTGGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1986  
QY 1561 ACACGCGCCACCGGAGAGGAGGAGTAAAGCTCTCACTTCAATGAGACCGTGGTGGTGGTGG 1620  
DB 1987 ACACGCGCCACCGGAGAGGAGGAGTAAAGCTCTCACTTCAATGAGACCGTGGTGGTGG 2046  
QY 1621 AGCAATTAAGCCGAGCTCTCTGCTCGGTGCAATGGTCTCGGCGCAGAGAGTGGAT- CCT 1679  
DB 2047 AGCAATTAAGCCGAGCTCTCTGCTCGGTGCAATGGTCTCGGCGCAGAGAGTGGAT- CCT 2106  
QY 1680 AGGTGTAAGCGCGGAGAGCGGAGTCTCAACATCGAGAACATCGTCAACACGAGTGCAC 1739  
DB 2107 AGGTGTAAGCGCGGAGAGCGGAGTCTCAACATCGAGAACATCGTCAACACGAGTGCAC 2166  
QY 1740 GCACGCTCATCTCGGAGACAGATTCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1799  
DB 2167 GCACGCTCAT- CCTCGGACAAAGATTCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2225











QY 767 GGTGCTGATGCTGGTCTGATGCTGGTGCTGGTCTGGTGGCAATAGTCTAAACAAGGCC 826  
Db 16814 AGTGTGGCGATGGTGGTGGTGGTGGTGGCAGCAGTCGTCAGCATCTGTTGTTATCAAGGCC 16873

QY 827 T 827  
Db 16874 T 16874

RESULT 11  
US-09-949-016-103758  
; Sequence 103758, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 103758  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-103758

Query Match 5.5%; Score 128.2; DB 3; Length 522;  
Best Local Similarity 59.6%; Pred. No. 4.3e-22;  
Matches 214; Conservative 1; Mismatches 144; Indels 0; Gaps 0;

QY 448 TGTCGTTGGGCCTGTTGTTGGTGTCTAGTAGCCGGTGGTGGCAGTGGTGCTG 507  
Db 18 TGATGGTGGTGATGGTGGTCTCGTGGTGGTGTGTTGTTGGTGGATGGTGAATGGTGA 77

QY 508 TGCCAAGGCTGGTCTGCTAGTGGTACCGTTGGGGCGTCTGTGCAGCCAGTAGACGGTCT 567  
Db 78 TGGTGGGGGTGGTGCTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 137

QY 568 CCCTGTTGTTGGGCCCTGTTTGGTGGTCTCTAGGTGGCCGTGGTGGCAGTGGTG 627  
Db 138 TTGTTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 197

QY 628 TGGTCCCAAGGCTGGTGGTGGTGGTACCCTAAGCCCGCTGACGGTCCAGTGGACGG 687  
Db 198 TTATGGTGATGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 257

QY 688 GTTTCCTCTCGTTGGAGACTTGCTGAAGGTCTACTGAGGTCTTTCTAGGTGGTGATGC 747  
Db 258 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 317

QY 748 TGGTTCCTGCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806  
Db 318 TGGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 376

RESULT 12  
US-09-949-016-14613  
; Sequence 14613, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307

[illegible]

Qy	499	TGCTGCTGGTCCCAAGGGTGGTGTAGTGTACCGTTGGGCGTCTGTGGCAGCCAGT	558
Db	243341	AGTTGCTCGTGGTGGAGGTGCTCGTGTGGT-GAGGTGCTCGTGGTGGAGGTTT	243283
Qy	559	AGACGGTCTCCCTGTTGTTGGGCTGTTGTTGGTGGTGTCTAGGTGGCGGTGGTGTGG	618
Db	243282	TCATGGTGCAGGTGCTCCTCGTGGAGATGCTCGTGGTGGTGGAGGTGCTTGTGATGATGG	243223
Qy	619	CAGTGGTGTGCTGGTCCCAAGGGTGGTGTAGTGTAGTGTACCCCTAAGCGCCGTGACCGTCC	678
Db	243222	AGGTGCTCCTGGTGGAGGTGCTCCTGTTGGAGGTGCTCGTGGAGGTGCTCGTGG	243163
Qy	679	AGTGGACGGTGTTCCTGCTCGTGGAGAGCTTGTCTGAAGGTGCTACTGGAGGTCTTCTAGG	738
Db	243162	TGGTGGAGATGCTCGTGGTGTGATGGAGGTGCTCGTGGTGGAGGTGCTCGTGGTGGAGG	243103
Qy	739	TGGTGTGCTGGTTCCTGCTGATGCTGCTGGTGTGCTGATGCTGGTGTGCTGCTGG	798
Db	243102	TTCTCATGGTGCAGGTGCTCCTGGTGGAGATGCTCTTGGTGGAGATGCTCGTGGTGA	243043
Qy	799	TGCTGGTG	806
Db	243042	TGGAGGTG	243035

Search completed: January 9, 2006, 00:27:13  
Job time : 304 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: January 9, 2006, 02:50:35 ; Search time 24 Seconds  
(without alignments)  
3026.819 Million cell updates/sec

Title: US-10-723-807-2\_COPY\_17\_771  
Perfect score: 4033  
Sequence: 1 LDVPRKPHAPTGEKRLTF.....EKNROTQVLHWHQTKKDL EE 755

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1162	28.8	793	T41703	dipeptidyl aminope
2	1054	26.1	818	A30107	dipeptidyl aminope
3	981.5	24.3	766	CDH026	dipeptidyl-peptida
4	981	24.3	760	S23752	dipeptidyl-peptida
5	949.5	23.5	792	A39914	dipeptidyl-peptida
6	902.5	22.4	931	A49737	dipeptidyl aminope
7	848.5	21.0	759	I38593	fibroblast activat
8	811.5	20.1	711	S66261	X-Pro dipeptidyl-p
9	779.5	19.3	803	I68600	dipeptidyl aminope
10	779.5	19.3	865	I54331	dipeptidyl aminope
11	773.5	19.2	803	A41793	dipeptidyl aminope
12	747	18.5	738	A87516	dipeptidyl peptida
13	708	17.6	741	JC5142	X-Pro dipeptidyl-p
14	584	14.5	743	T37700	probable dipeptid
15	545	13.5	779	T25173	hypothetical prote
16	545	13.5	799	T25174	hypothetical prote
17	508	12.6	829	T19514	hypothetical prote
18	430.5	10.7	931	T32919	hypothetical prote
19	344.5	8.5	795	F82858	dipeptidyl-peptida
20	243	6.0	657	F70025	probable acylamino
21	210.5	5.2	631	H75007	probable acylamino
22	189.5	4.7	618	D87651	prolyl oligopeptid
23	189.5	4.7	674	B84381	acylaminoacyl-pept
24	187.5	4.6	591	H72474	probable acylamino
25	186	4.6	598	F84199	hypothetical prote
26	186	4.6	642	C71137	hypothetical prote
27	184	4.6	721	T09631	probable acylamino
28	179.5	4.5	709	B82580	alanyl dipeptidyl
29	171	4.2	683	E87495	prolyl oligopeptid

ALIGNMENTS

RESULT 1

T41703  
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T41703  
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z22011  
A:Accession: T41703  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-793 <MUR>  
A:Cross-references: UNIPROT:O14073; UNIPARC:UPI000013AB95; EMBL:AL031180; PIDN:CAA20138  
A:Experimental source: strain 97Zh-; cosmid c2E11 -chimeric  
C:Genetics:  
A:Gene: SPAC2E11.08  
A:Map position: 1  
C:Superfamily: dipeptidyl-peptidase IV

Query Match 28.8%; Score 1162; DB 2; Length 793;  
Best Local Similarity 36.1%; Pred. No. 1.7e-67;  
Matches 272; Conservative 142; Mismatches 294; Indels 46; Gaps 23;  
QY 22 ETVVKQAITPTSGRSVQWL--SGAEDGSYYVAAEDGSLTIENIVTNESTRILIP-ADKIPTG 78  
DB 58 DDIVLQKYKPSYKQVNWIDSQLKD---TFLVKYGDLLINIQDPYNLTKLFSVSDLVYNG 114  
QY 79 --KEAFNYWIHPDLSVLWASNHTKQYRHSFFADYYVQDVESLKSVPML--PDQEGDIQY 134  
DB 115 IQLDYDSYISFDKAYLVSVNKSQRWRHSSPAQYYLYNTET-KDYNMLQGDNEHWTISL 173  
QY 135 AQWSPGVNITAFVRENDLYVWN-GVTRITDDGGDFMFGVDPDWIYEIEILGDRVALWP 193  
DB 174 AEWSPGTGHQLSFYNNNDLYVRKNDGNVQLTYDGTVDVFNGLTMDIYEIEVLSVSPSTIMW 233  
QY 194 SPDGEVLAYLSFNETGVTYVQYVYNDNOEIAPAYP-----WELKIRYKPVKSVQTNVTIL 248  
DB 234 SPDSDKIAFLKUNSEIPIY--HYPLYTAELDPSLPEFDYKDKMAIKYKPGNPNFSVSL 291  
QY 249 SLINIASKEVKQAPIDAFES-TDLIIIGEVAWLTDTHTTVAAKAFNRVQDQKVVAVDTAS 307  
DB 292 FVADLNSNASSNFSLWHEPLAEPPVQVNLVWNT--SSVLVQPTNRNSTCITARLLDTEL 349  
QY 308 NKATVLSDRDGTGWLNDLLSMKYIGPIKPS-----DKDAYIIDSDHSGWAHLIYLPVS 362  
DB 350 KSIHTVKTECLEBEGWYEYVQSSAKWF-PLNNSLVWENWSDGYF-DIALDDYNNHLAFIPFN 407  
QY 363 GGEPIPLTKGDMVET-SILSIDOERQLVYVYLSQHHSTERHLYSVSVYFAVTPLVDD-T 420  
DB 408 GSSPIYLTSGAMVDTPGIHIDGDFGNVFLATLKDSTERHLYSVSLDTLEIYGITDNGE 467

30 168 4.2 606 2 T35378 probable peptide h  
31 168 4.2 676 2 C97775 acylamino-acid-rel  
32 165.5 4.1 732 1 S07624 acylaminoacyl-pept  
33 164.5 4.1 632 2 E75057 peptidase PAB1418  
34 162.5 4.0 622 2 F71174 hypothetical prote  
35 161 4.0 654 2 AD3183 peptidase limporte  
36 160 4.0 637 2 S75772 hypothetical prote  
37 156 3.9 729 2 A97747 oligopeptidase B (  
38 152 3.8 732 1 JC4655 acylaminoacyl-pept  
39 149.5 3.7 629 2 T15945 hypothetical prote  
40 149.5 3.7 659 2 F72568 probable acylamino  
41 149.5 3.7 955 2 F84914 hypothetical prote  
42 146.5 3.6 655 2 E75551 probable acyl-pept  
43 141.5 3.5 764 2 JC8016 acylaminoacyl-pept  
44 140 3.5 13055 2 T16580 hypothetical prote  
45 139.5 3.5 732 1 JU0132 acylaminoacyl-pept

```
Qy 421 VAAWWSAFSANSYIILTYGGDPVYPQELYYTNTSKPLRTITDIAKVLQIKDYALPNI 480
Db 468 DEGYSTSFSPFGDFVYLVNHPDPVPMQELASTKDKOYCLISLETNSLKKOQLSITLPSV 527
Qy 481 TYFELPLSPSGTLLNMQRLLPGFSPDKKYPILFTYPCGPGAOEVTKERQALNPKAYVASD 540
Db 528 EYGLKTF-NDTTFMFERRPRNFVNKKYPVLFFAYGPGGQQVAKLFR-VDFQAYLASH 585
Qy 541 SELEYVVTWVDNRGTGFGKGRFSRAVTRQLGLLEAEDQIYAAQQAANIPIWIDADHIGIW 600
Db 586 PDFFIVVLLDGRGTGFGNAFRYSVRHLGEWESYDQGAQKFWADLPFDVDENHVGIWG 645
Qy 601 WSPGGLYTSKVLKDSGAFTLGVTAPVSDWRFYDSMTYRYMYTKLSTNBERGYTSAVRK 660
Db 646 WSYGGYLTKLTKETQD-VFSYGMVAPVTDWRLYDSVYTERYMDLPQYNKEGYKNSQIHD 704
Qy 661 TDGKFNVEGGFLIOHGTGDDNVHQNLSAALVDLLMGDGVSPKLSHS-----QWFTSDHGIS 717
Db 705 YEKFKQLK-REFVAHGTGDDNVHFGHSMHLM-----DGLNLANCYNIDMAVFPDPSAHSIS 758
Qy 718 YHGGGVFLYKQARKLYOEKNR---QTQVLMHQW 748
Db 759 YHNASLSIYHRLSEWIGDALGRIDPSTGVRQHRW 792

RESULT 2
A30107
Dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YHR028C
C;Species: Saccharomyces cerevisiae
C;Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 09-Jul-2004
C;Accession: S46780; A30107
R;Du, Z.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 8082.
A;Reference number: S46773
A;Accession: S46780
A;Molecule type: DNA
A;Residues: 1-818 <DUZ>
A;Cross-references: UNIPROT:P18962; UNIPARC:UPI0000031A5F; EMBL:U10399; NID:g500689; PID
R;Roberts, C.J.; Pohlrig, G.; Rochman, J.H.; Stevens, T.H.
J. Cell Biol. 108, 1363-1373, 1989
A;Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an in
F;30-45/Domain: transmembrane #status predicted <TM>
A;Reference number: A30107; MUID:89174971; PMID:2647766
A;Accession: A30107
A;Molecule type: DNA
A;Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRRLET', 189-199, 'N', 201-365, 'DFKRGKERKF', 376-57
A;Cross-references: UNIPARC:UPI0000172A3C; EMBL:X15484
A;Note: the authors translated the codon ACC for residue 572 as Asn
C;Genetics:
A;Gene: SGD:DAP2; STE13; MIPS:YHR028C
A;Cross-references: SGD:S0001070; MIPS:YHR028C
A;Map position: 8R
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
F;30-45/Domain: transmembrane #status predicted <TM>
F;63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 26.11%; Score 1054; DB 1; Length 818;
Best Local Similarity 33.11%; Pred. No. 1.9e-60;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNVTVVKQAITPTSRSVOWLS-----GARDGSYV 48
Db 48 IPHSHNTPDYQEPNSNYNDGKLKVSFS-VVRNNTFFQPKYHELOWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNTNESRTLIPADKIPTKGEAFNWIH-----PDLSSVLW 95
Db 107 TFMND-SYVVKSVYDSDSVNLLEGKT-----FIHQNLQAVESITASPDKLRLI 156
Qy 96 ASNHHTKQYRHSFADYYVDVLSKSVPLMPDQEGDIQVAAQSPVGNITAFVRENDLYVW 155
Db 157 RTNSVQNRHSHSTFGSYFVVD-----KSSSSFEIGNEVALAIWSPNSNDIAYQDNNIYY 212
```

```
Qy 156 ---DNGTIVTITDDGGPMDFHGVDPDWIYEEILGDRVALWFSPDGEYLAVLSFNETGVPT 212
Db 213 SAJSKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWWSPTGDTLAFLLKIDSEVGE 272
Qy 213 YTVQYVMDNOEIAIPAPFWELKIRYPKVQSQTNPVTLSLLNIASKEVKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIPEMRSIKYPKSGTNPFAELWVYSMKDGTSPHPRISGNKDGSL 330
Qy 272 IIGEVAWLTDTHTTAAKAFNRVQDQOKVAVDTASNKAIVISDRDGTDCWLDNLLSMKY 331
Db 331 LITEVTWVGNGN--VLVKTTDRGSDILTVELIDTIAKTSNVVRNNESSNGGWWEITHNTLF 388
Qy 332 IGPIKSPDKDAY-YIDISDSHGAHLVLPVSGEPI-PLTKGDWEYTS-ILSDQERQL 388
Db 389 IPANETDFRPHNGVDILPIGGYHNHAYFENSNSHYKTLITTEGKWEVVGNGLAFDSMNR 448
Qy 389 VYILSTOHSTERHLXSVSY-STFAVTPLVDDTVAAYWSASFANSYGYIILTYGGDPVYP 447
Db 449 LYFISTKRSSTERHVYIIDLRSPNEIIEVDTSDGVDYVDSFGSGRRFGLITYKGPVYP 508
Qy 448 QELYTTNST-----KPLRTITDIAKVLQIKDYALPNIYFEIPLPSE---- 491
Db 509 QKIVDFHSRAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGSP--DKKYPILFTYPCGPGAOEVTKERQALNPKAYVASDSELYVTW 548
Db 569 ILVNSYILPNDFDETLSDHYPVFFFAYGPGNSQOVVKTF-SVGFNEVVA--SQLNALIV 625
Qy 549 TVDMNRGTGKGRFSRAVTRQLGLEAEDQIYAAQQAANIPIWIDADHIGIWMGSGFGYLT 608
Db 626 VVDRGRTGFGKQDFRSLVRDLGDEYARDQISAASLYGSLTFVDPQKLSLFGWSYGGYLT 685
Qy 609 SKVLKDSGA-FTLGVTITAPVSDWRFYDSMYTERYMYTKLSTNBERGYTSAVRKTDGKNV 667
Db 686 LKTLKDGGRHFKYGMVAPVTDWRFYDSVYTERYMYTHPTQENPFDGYVESSVHNVTALQA 745
Qy 668 EGGFLIOHGTGDDNVHQNLSAALVDLLMGDGVSPKLSHSQWFTSDHGISYHGGGVFLYK 727
Db 746 -NRELLMHGTGDDNVHFNQSLKFLDLDLLNGVENYDVHV--FPDSHHSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 3
CDHU26
dipeptidyl-peptidase IV (EC 3.4.14.5) - human
N;Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 23-Aug-1996 #text_change 09-Jul-2004
C;Accession: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520
R;Misumi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
Biochim. Biophys. Acta 1131, 333-336, 1992
A;Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a ser
A;Reference number: S24313; MUID:92329551; PMID:1352704
A;Accession: S24313
A;Molecule type: mRNA
A;Residues: 1-6, 'I', 8-766 <MIS>
A;Cross-references: UNIPROT:P27487; UNIPARC:UPI000016A880; EMBL:X60708; NID:g35335; PIDN
R;Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sapin, C.; Trotot, P.; Barbat, A
J. Biol. Chem. 267, 4824-4833, 1992
A;Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer
IV mRNA levels during cell differentiation.
A;Reference number: A42408; MUID:92165847; PMID:1347043
A;Accession: B42408
A;Molecule type: mRNA
A;Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DAR1>
A;Cross-references: UNIPARC:UPI0000052ACB; GB:M80536; NID:g181569; PIDN:AAAS2308.1; PID:
A;Experimental source: intestine
A;Note: this sequence corresponds with the author's translation
A;Accession: A42408
```

A:Molecule type: mRNA  
A:Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-711,'G',713-766 <DAR2>  
A:Cross-references: UNIPARC:UPI0000172A2B; GB:M80536; NID:g181569  
A:Note: sequence extracted from NCBI backbone (NCBIN:83986, NCBI:P:83988); this sequence  
R:Corvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Gastroenterology 101, 618-625, 1991  
A:Title: Expression of sucrose-isomaltase and dipeptidyl-peptidase IV in human small intestine  
A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: B61136  
A:Molecule type: protein  
A:Residues: 1-15,'X',17-22 <GOR>  
A:Cross-references: UNIPARC:UPI0000172A2C  
R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.  
Biochem. J. 311, 835-843, 1995  
A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a  
A:Reference number: S59510; MUID:96067599; PMID:7487939  
A:Accession: S59510  
A:Molecule type: DNA  
A:Residues: 1-31 <BOE>  
A:Cross-references: UNIPARC:UPI000016B4A6; GB:S79876; NID:g1195574; PIDN:AA35614.1; PID  
R:Fanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H.  
J. Immunol. 149, 481-486, 1992  
A:Title: Cloning and functional expression of the T cell activation antigen CD26.  
A:Reference number: I56154; MUID:92325476; PMID:1352530  
A:Accession: I56154  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-436,'S',438-766 <TAN>  
A:Cross-references: UNIPARC:UPI000004F7BF; GB:M74777; NID:g180082; PIDN:AA51943.1; PID:  
R:Abbott, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.  
Immunogenetics 40, 331-338, 1994  
A:Title: Genomic organization, exact localization, and tissue expression of the human CD  
A:Reference number: S59857; MUID:95012454; PMID:7927537  
A:Accession: S59857  
A:Molecule type: DNA  
A:Residues: 1-436,'S',438-766 <ABB>  
A:Cross-references: UNIPARC:UPI000004F7BF; EMBL:U13734  
C:Genetics:  
A:Gene: GDB:PPP4  
A:Cross-references: GDB:125239; OMIM:102720  
A:Map position: 2q24.3-2q24.3  
A:Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3  
C:Superfamily: dipeptidyl-peptidase IV  
C:Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; homodimer; proteinase; transmembr  
F:1-6/Domain: intracellular #status predicted <INT>  
F:7-28/Domain: transmembrane #status predicted <TN>  
F:29-766/Domain: extracellular #status predicted <EXT>  
F:85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #status  
F:630,708,740/Active site: Ser, Asp, His #status predicted

Query Match 24.3%; Score 981.5; DB 1; Length 766;  
Best Local Similarity 32.5%; Pred. No. 8.7e-56;  
Matches 249; Conservative 131; Mismatches 304; Indels 83; Gaps 27;

QY 11 TGGSKRLTNETVTKQAITPTSRVOMLSGADGSGVVAAGDGLTIENIVNESRTLI 70  
DB 36 TADSRKTYLTLD-YLNKTYRLKLYSLRWIS---DHELYLKQENNLVFNAYGNSSVFLE 91  
QY 71 PADKIPGKEAFNVNHPDLSSVLWASNHTKQYRHSFFADYVQDVESLKSVPIMPDOE- 129  
DB 92 NSTFDFEGHINDYISPDQGFILLENNYKQWRHSYASY---DIYDLNKRQLITEERI 148  
QY 130 -GDIQAQMSFVGNITAFVRENLDY--WDNGTVTRITDDGGPD-MFHGVPDMDIYEEIL 185  
DB 149 PNNTQWTVWSPFGHKLAVYVNDIYKIEPNLPSYRITWTGKEDIYNGITDMWYEEVF 208  
QY 186 GDRYALWSPDGEVLAFLSNETGCVPTTYVQYVNDQGEIAPAYPWEKIRYPKVQSNPT 245  
DB 209 SAYSLAWSPNGTFLAYAQFNDTEVPLIEYSFSDS---SLQVPKTVRVVPYKAGAVNP 265  
QY 246 VTLSLNI--ASKEVKQAPIDAFESTOLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297  
DB 266 VKFVVNTDLSVSTNATSIQTAPASMLGDHVLCDVTWATQER--ISLOWLRRIQNS 323

QY 298 -QKVAVAVDTASNKATVISDRD-----GTDGWLNDLLSMKYIGPIKPS-----DKDAYYTD 346  
DB 324 VMDICDYDESSGRWNCCLVARQHIESTTGW-----VGRFRPSEPHFTLDGNSFYKI 374  
QY 347 ISDHGSAWHLVLPVPSGGSEPTPLTKGDWEVTSILSIDQERQLVYVLSQHHSTE--RHLY 404  
DB 375 ISNEEGYRHCYFQIDKKDCTFITKGTWEVIGIEALTSD--YLYVISNEYKMGPPGGRNLY 432  
QY 405 SVSYSTEA-VTPEL---VDDTVAAYWSASFSANSYGVIITYGDPVYQBYQLYTTNSTKPLR 460  
DB 433 KQLLDYTKVTCLSCELANPERQYYSVSFSKAKYQLRCSGPGPLPLYTLHSSVNDKGLR 492  
QY 461 TITDNA---KVLQEIK-----DYALPNITYPELPLPSETLVNQWRLPPGSPDKKYP 511  
DB 493 VLEDNSALDKMLQNVQMPKSLDFILNETKEWYQW-----ILPHFDKSKCYPL 542  
QY 512 LFTPYGGGCAQVTRKQALNFKAVASDSLEYVTWTVDNRGTGFGKGRKFSASVTRQLG 571  
DB 543 LLDVYAGPCSQKADTVFR-LNWATYLASTENI--IVASFDGRSGYQGDKIMHAINRRLLG 599  
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGWGSFGYLTSLKVLBKDSGAFTLGVITAPVSDW 631  
DB 600 TFEVSDQIEAAQFQKMGFVDNKRITAIWGSYGGVTSVNLGSGSGVFKCGIATAVPVSRW 659  
QY 632 REYDSNMYTERYM--KTLSTNEEGYETSAY-RKTDGPKNYEGGFLIOHGTGDDNVHFNQNSA 688  
DB 660 EYDVSYYTERYMGLPPTPDNLHYRNTVMSRAENFKQVE--YLLIHGTADDNVHFPQSA 717  
QY 689 ALVDLLMGDGVSEPKLHSONFTDSDHGISYHGSGVFLYKQLARKLYQ 735  
DB 718 QISKALVDVGDV---FQAMWYTDDEHGIASSTAHQHIYTHMSHFIKQ 761

## RESULT 4

dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse

N:Alternate names: CD26 alpha subunit; THAM alpha subunit

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S23752; A46465; A56030

R:Marguet, D.; Bernard, A.M.; Vivier, I.; Darnoul, D.; Naquet, P.; Pierres, M.

J. Biol. Chem. 267, 2200-2208, 1992

A:Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-d

A:Reference number: S23752; MUID:92129288; PMID:1370813

A:Accession: S23752

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-760 &lt;MAR&gt;

A:Cross-references: UNIPARC:UPI0000172A2D; EMBL:X58384

R:Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;

J. Immunol. 147, 447-454, 1991

A:Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase

A:Reference number: A46465; MUID:91302787; PMID:1712807

A:Accession: A46465

A&gt;Status: preliminary

A:Molecule type: protein

A:Residues: 1-20 &lt;VIV&gt;

A:Cross-references: UNIPARC:UPI0000172A2E

A:Experimental source: M14.T thymoma cells, Swiss nu/nu

A:Note: sequence extracted from NCBI backbone (NCBI:P:42236)

R:Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.

Biochemistry 33, 15204-15214, 1994

A:Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.

A:Reference number: A56030; MUID:95092780; PMID:7999781

A:Accession: A56030

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 746-760 &lt;BER&gt;

A:Cross-references: UNIPARC:UPI0000172A2F; GB:U12620

C:Genetics:

A:Gene: CD26

C:Superfamily: dipeptidyl-peptidase IV







Query Match	21.0%;	Score	848.5;	DB	2;	Length	759;
Best Local Similarity	30.0%;	Pred.	No. 3.8e-47;				
Matches	229;	Conservative	139;	Mismatches	291;	Indels	105;
Gaps	29;						
QY	4	PRKPHATGEGSKELT-----FNEVVVQQAITPTSRSVQVLWSGAEDGSGVYVAAEDGSLTIE	59				
DB	28	PSRVHNSSEENTMRALTLDKILNGTFSYKTFPP-----NWISGQE--YLHQSDANNIVLY	79				
QY	60	NIVNSESRTLIPADKIPGKEAFNYIHPDLSSVLMASNNHTKYRHSFFADYVYVQDVESL	119				
DB	80	NIETGQSVYIIL-SNRMTKSVSNASNYGLSPDRQFVYLESYDSKLRYSVTATYIYDLSNG	138				
QY	120	KSVPLMPDQEG-----IQYQAWSPGVNNTIAFVRENDLIYWD--NGIVTRITDDGGPD-M	171				
DB	139	EFV-----RGNELPRPIQYLCWSPVSGKLAYVYQNNIYILKQRPDGPFPQITFNGRENKI	192				
QY	172	FHGVPDWIYBEIILGDRVALMFWSPDGEVLAYLSFNETGVTYTVQYYMNDQEIAPAYPWE	231				
DB	193	ENGIQPDWYIEBMLPTKVALWMSPNKGLAFABFNKDQIDPIVAIYSYQDEQ-----YPT	247				
QY	232	LKIRYPKVSQNTPTVTLISLN-----TASKEVQKAPIDA-FESTDLIIGEVAMLTDTHT	284				
DB	248	INIPYKAGAKNVVRIIFIIDTYPAYVGPQEV---PVPAMIASDDYFSLTWYTDER-303					
QY	285	TVAAKAFNRVQDQKQVAVD-----TASNKATVJISDRDGTGWLNDLLSKMYIGPIKPS	338				
DB	304	-VCLQWLKEVQNVSVLISICDFREDQWTDCPKTQEHIBESRTGWAGGPFVSR---PVFSY	359				
QY	339	DKDAYVIDISDHSWAHLYLFPVSGEPIPLTKGDWEVTSILSIDOERQVLYLST--QH	396				
DB	360	DAISYKIPSDKDGKXKHIYIKDTVENAFOITSGKWEAINIPRVQDS--LFYSNFEFE	417				
QY	397	HSTERHLXSVSYSTR-----AVTPIVDDTVAAIYWSASFANSFGSYILTYGGPDVPQBEL	451				

```

C;keywords: dipeptidipeptide nyrarolase

Query Match      20.1%; Score 811.5; DB 2; Length 711;
Best Local Similarity 30.0%; Pred. No. 8.7e-45;
Matches 207; Conservative 135; Mismatches 278; Indels 71; Gaps 20;

149  ENDLYVMD--NGVTRITDDGGP--DMFHPVDWIIYEEILGDRYALWFSPPGEVLYLSF 205
144  DNNLYQDLNTGKITQITTDGKNEIINGLDWVYEEEFGHADYYQW--NKGADALVVFRR 202
206  NETGVPVTVQYYMNDQEIAPAYPWELKIRYPKVSOTNPVTTLTLLNIIASKEVQKAPIDA 265
203  DERKVPENIPIYYQN-----LYPKLMTYKYPKAGENSAVTAYLYQLSSG--KSAQLNF 255
266  FESTDLIIIGEVAWLTDHTTVAAKAFNRVQDQKQVAVDVTASNKATVJSD--RDGTDGWL 323
256  GSSEKIYIPQL-FQTNANDEIVVATNRHQNKVDLLKYN-----KTAAVSKLFTETDNAMI 311
324  --DNLLSMKIPIPKPSDKAYIIDISPHSGWAHLYLPFPVSGGEPIPLTKGDWEVTSILS 381
312  ETDN-LTMEFL-----DDNSFLWASERDGRHLYWYDAAGKLKKQVSKGDWEIINYVG 363
382  IDQERQLVYTLSTQHSHTERHLYSVSYSTFAVTLPLVDDTVAAYWASGSFANSGYILTGV 441
364  YNPKTKEVYIQTEKGSINKVVSUKLINFT-GKTQLLSNAEGNN-SNAFSKTFNYFINTSS 421

```



Db 542 SFSHSMDFLLKCRGPGVPMVTVENTTDDKKQFDLETNEHVKKAINDRQMPKVEYRDIEI 601  
Qy 488 PSGETLAV-MORLPPG-FSPDKKYPILFTPYGGGCAQEVTKR-----WQALNFKAYVASDS 541  
Db 602 ---DDYNLPMQILXPAITFTDTHYPLLLVVDGTPGSGQVAEKFEVSMETVMVSSHGA--- 655  
Qy 542 ELEVVYTWVNRGTGFKGRKFRSAVTRQGLLEAEQDIYAAQQAANIPIWDADHIGW 601  
Db 656 ---VVVKCDGRSGFGQTKLLHEVRRRLGLEEKQMEAVRTMLKEQYIDRTVAVFGK 711  
Qy 602 SFGGVLTSKVL-----EKDSGAFTLGVITAPVSDMRFRYDSMYTERVMKTLSTNEBGE-TS 656  
Db 712 DYGGYLSYIILPAKGENGGQITFCGSAISPITDFKLYASAFSERYLGLHGLDNRAYETK 771  
Qy 657 AVRTKDGFKNVEGFLIOHGTGDDNVHFNQSAALV-DLLMGDG-----VSPEKLH 710  
Db 772 VAHRVSALE--EQQLIITHPTADEKIHQHTAELITQLIRKANYSLQIYPDESH--YFT 827  
Qy 711 DSDHGISVHGGVFYKQLAKRLQVQKRNQRTQVLMHQWTKDLBE 755  
Db 828 SS--SLKQH-----LYRSII--NFFVECFRIQDKLPTVTVAKEDEE 864

RESULT 11  
A41793  
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41793  
R:Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimaeaki, S.  
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992  
A:Title: Differential expression of two distinct forms of mRNA encoding members of a dipeptidyl aminopeptidase family  
A:Reference number: A41793; MUID:92108018; PMID:1729689  
A:Accession: A41793  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-803 <WAD>  
A:Cross-references: UNIPROT:P42659; UNIPARC:UPI000002A83C; GB:M76429; NID:9408719; PIDN:  
A>Note: sequence extracted from NCBI backbone (NCBIF:75138)  
C:Superfamily: dipeptidyl-peptidase IV  
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein  
F:257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.2%; Score 773.5; DB 2; Length 803;  
Best Local Similarity 29.3%; Pred. No. 3.1e-42;  
Matches 210; Conservative 124; Mismatches 295; Indels 87; Gaps 26;

Qy 37 QWLSGAEDGSYVYAAEDGSLTIENIVTNESTRLLIPADKIPTGKEAFNYWIHPDLSSVLWA 96  
Db 87 KWSI---DKETIYREQGSSVILRNVTNTSTVLEGGKIES-LRAIRYEISPDREYALFS 142  
Qy 97 SNHTKQYRHSFPADYVQDVESLKSVPMLPDQEGD-----IQVAQWSPVGNITIAFVR 148  
Db 143 YNVEPIYQHS-VTGIVY-----LSKIPHGDPQSLDPPEVSNAKLQYAGNGPKGQQLIFIF 196  
Qy 149 ENDLYWMD--GTVTRITDDGGD-MFHGVPDWIYEEIEILGDRYALWSPDGEYLAISLF 205  
Db 197 ENNIYCAHVQKQAIRVVSTGKEGVYNGLSDWLEEEILKTHIAHWSPDGTRLAYATI 256  
Qy 206 NETGVTPTYVQYMDNQETAPAYPWLKIRYPKVQTNPTVTLSL--LNIAKEVKQAPI 263  
Db 257 NDSRVPVMEPLTYTGS-----VYPTAKPHYHYKAGCENPSISLHVIGLNGPHTDLEMTPP 311  
Qy 264 DAFESTDLIIGEVAMLTDTHTTAAKAFNRVQDQKQVAVDVTASNKATVISDRDGTGWL 323  
Db 312 DPMREYIITWVKWATS--TKVAVNWLSPAQNVSILTCDATGTGCTK-KHEDESEAWL 368  
Qy 324 DNLLSMKYIGPKPSD-KDAYIYIDISDHGWAHLVFPVSGGEP-----IPLTKGDWE 375  
Db 369 HR----QNEEPVSKDGRKFFFRAPQGGQKGFYHITVSSSQPNSSNDNIQSITSGDW 424  
Qy 376 VTSILSIDERQLVYLSLTHQHSHTERHLYSVS-YSTFATVPLVDVTV--AAYWSASFSA 432

Db 425 VTKILSYDEKRSQIYPLUSTEDLPERRQLYSASTVGSFNRQCLSCDLVDNCTYTSASFSPG 484  
Qy 433 SGYILTYGGDPVPYQELYTNSTKPLRTITDQNAKVLQIKDYALPNITY-----FEL 485  
Db 485 ADFELLKCEGPGVPTSVHNTTDDKKQFMLETNEHVQKAIASDRQMPKVEYRKIETDDYNL 544  
Qy 486 PLPSGETLNVORLPPG-FSPDKKYPILFTPYGGGCAQEVTKR-----WQALNFKAYVASD 540  
Db 545 PI-----QILKPAITFTDTHYPLLLVVDGTPGSGQVAEKFAVWTETVMVSSHGA-- 593  
Qy 541 SELEYVTWTVNRGTGFKGRKFRSAVTRQGLLEAEQDIYAAQQAANIPIWDADHIGW 600  
Db 594 -----VVVKCDGRSGFGQTKLLHEVRRRLGLEEKQMEAVRVMLEKPYIDKTRVAVFG 648  
Qy 601 WSPGGYLTSKVL-----EKDSGAFTLGVITAPVSDMRFRYDSMYTERVMKTLSTNEBGEYS 656  
Db 649 KDYGYLSYIILPAKGDQGPVFCGSAISPITDFKLYASAFSERYLGLHGLDNRAYEMA 708  
Qy 657 AVRTKDGFKNVEG-FGLIOHGTGDDNVHFNQSAALV-DLLMGDG-----VSPEKLH 705  
Db 709 KV--AHRVSALEGGQFLVIHATADEKIHQHTAELITQLIRKANYSLQIYPDESH 762

RESULT 12  
A87516  
dipeptidyl peptidase IV [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: A87516  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 99, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-738 <STO>  
A:Cross-references: UNIPROT:Q9AGE0; UNIPARC:UPI000000C7616; GB:AE005673; NID:913423647; P:  
C:Genetics:  
A:Gene: CC2154

Query Match 18.5%; Score 747; DB 2; Length 738;  
Best Local Similarity 29.5%; Pred. No. 1.4e-40;  
Matches 214; Conservative 116; Mismatches 292; Indels 104; Gaps 24;

Qy 28 AITPTSRQVQWLSGAEDGSYVYAAEDGSLTIENIVTNESTRLLIPADKIPTGKEAFNYWIH 87  
Db 44 ALSPPGKRVYILKGKPE-----AANIQDLWAADVKGGEPRYLIDSAAALSSGDK----- 91  
Qy 88 PDLSSVLWASNHTKQY-----RHSPF-----ADYVQDVESLKSVPMLPDQEG 130  
Db 92 -ELSEAKARERARVARSARGIVEYSDWQRGRFILVPLDGLDLYLDAVADGK-ITRLTETPG 149  
Qy 131 DIQYQWSPVGNITIAFVRENLYWMD--NGTVTRITDDGGDPMFHGVPDWIYEEIEILGDR 188  
Db 150 DEVDKAVSPKGGYSVYRDQNLITKPVAGGAETALTDTGKDALSFGEVAFVQEBEL--DR 207  
Qy 189 Y-ALWFSPDGEYLAISLSPNETGVPPTYVQYMDNQEIAPAYPWLKIRYPKVQTNPTVT 247  
Db 208 FGYQWSPDESRIYVTRVDESGVD-----IVPRADIGPGGATVNVQRYPRAGRPAVVD 261  
Qy 248 LSLNIAKEVKQAPIDAFESTDLIIGEVAMLTDTHTTAAKAFNRVQDQKQVAVDVTAS 307  
Db 262 LFRVRLASGV--TALDLGANKDIYVARVAMSAD-GKTVYVQRLSRDQKTLDLAPDAAT 318  
Qy 308 NKA-TVISDRDG-----TDG---WLDNLLSMKYIGPKPSDKDAYIIDSHS 351  
Db 319 GAGKTILDTDDPHFEVSNDFRPLTDGTFW-----GSEKD----- 354  
Qy 352 GWAHLVLPVSGGEPILTKGDWEVTSILSIDERQLVYLSLTHQHSHTERHLYSVSYSTF 411

```
Db 355 GNQHLYRYAADGKLIQAITKGDWPIVGLGVDEARKAIFAISIDTPIERRLYEVSYAK- 413
QY 412 AVTPLVDDTAAAYWSAFSANSYIILTYGGDPVYQELTYTNSTKPLRTITDIAKVLKLE- 470
Db 414 POKPALTSAGGWAAKVADNGGAFAGTYSDPKTSQALYSADGRVRVTEEN-KLAEG 472
QY 471 -QIKDYA--LPNITYFELPLPSGETINVMQRLPPGSPDKYPILFTPYGPGGAQEVTKR 527
Db 473 HPYWPYAANLPQEPFGLSKAADGETLHYEILKPIGFDPAKKYPAIVSYVYGGPHQAQRVMKN 532
QY 528 WALNFKAYVASDSLEYVTVTDNRGTGFKGRKFRSAVTROLGLEAEQIYAAQQAAN 587
Db 533 WHSPSERTYL-----BAGYVIFKLDNRGSGNRSKAFMRALDRKLGTVESVDQLLGAKFLAS 588
QY 588 IPWIDADHIGIWGWSFGGYLTSSKLEKDSGAPTLGVIITAPVSDMRFYDSMYTERMYKTL 647
Db 589 QYVVDADKLGVMGWSYGGFMALMLITAEPTPKAGAAGAPPTENSLYDTATYTERMYGKPD 648
QY 648 TNEEGYETSAV-RKTDGFKNVGEGFLIQHGTGDDNVHFQNSAALVDLLMGDGV-----S 700
Db 649 ENKAGYAYSDINNRIID--KLAFGSLLLLHGMADDNVIPENSTRMAALQKAILPEMAMY 706
QY 701 PEKLS 706
Db 707 PGERHS 712

RESULT 13
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
C:Accession: JC5142
R:Kasahima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expression
A:Reference number: JC5142; MUID:971164011; PMID:9010758
A:Accession: JC5142
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-references: UNIPROT:P95782; UNIPARC:UPI00000B5F8A; DDBJ:D83263; NID:gl753196; PI
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl res
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted <TM>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted

Query Match 17.6%; Score 708; DB 2; Length 741;
Best Local Similarity 28.3%; Pred. No. 5e-38;
Matches 208; Conservative 132; Mismatches 303; Indels 92; Gaps 28;
QY 23 TVVVKQAITPSTRSVQWLSGAEDG-----SYVYAAEDGSLTIEN-IVTNESRTILIPADK 74
Db 38 TLMKPKVADGSRVFLRCKSDRQLDLWSYDGSQTRLLVDSKVLPGTETLSDEEK 97
QY 75 IPTGKEAF-----NYWTHPDLSSVLWASNHTKQYRHSFFADYVQDV--BSLSKSVPLM 125
Db 98 ARRRQRIAAMTGIVDYQWSPDAQRLPLPLG-----GELYLDLQEGKAARVL 147
QY 126 PQEGDIQYQAQSPVGNITAFVRENDLYWD--NGTVTITDDGGPDMEHGVDPDIYEE 183
Db 148 THGEGPATDAKLSPGGFFSIFGRNLWIDLASGRQMQLTADGSTTIGNGAIFVADSE 207
QY 184 ILGDY-ALWESPDPGEYLAYSFNETGVPTY-TVQYMDNQETAPAYPWEKIRYPKVSQ 241
Db 208 M--DRHTGVMWAPDDSAIYARIDESFPVQKRYEIVADRTDV-----IEQRPAAGD 258
QY 242 TNPVTLSLNTASKEVKQAPIDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQOK-- 299
Db 259 ANVQVKLGVISPAEQAOQW--IDLGKEQDIYLARVNRDPQHLSPQRS----RDQKKLD 313
```

```
QY 300 VVAVDTASNKATVISDRDGTGWLNDLSMKYIGIPKPSDKDAYIDISDSHGWAHLIYLF 359
Db 314 LVEVTLASNOQRVLA-HETSPVTWPLHNSLRF-----DDGSILWSSERTFGOHLVRI 365
QY 360 PVSGEPIPLPKGDWEVTSILSIDOEROLVYVYLSLTOHHSTERHLYSVSYSTFAVTPLVDD 419
Db 366 D-SKGKAAALTHGNWSVDLAVDEKAGLAYFRAGIESARBSQIYAVPLQ--GGQPQRLS 422
QY 420 TVAAWMSASFANSYIILTYGGDPVYQ-ELYTTNSTKPLRTITDIAKVLQIKDYALP 478
Db 423 KAPGHSASFARNASVYVDSHNSNSTPQIELFRANGEK-IATLVEN-----DLADPKHP 476
QY 479 NITYFELPLP-----SGET-INVMQRLPPGSPDKYPILFTPYGPGGAQEVTKRW 528
Db 477 YARYEAQRPEVFGTLTAADGKTPLYNSVIKPAQDPKRYPVAVVYVGGPASQTVTDSW 536
QY 529 QALN---FKAYVASDSLEYVTVTDNRGTGFKGRKFRSAVTROLGLEAEQIYAAQQA 585
Db 537 PGRGHLFNQYLAQGG---YVFSLDNRGTPRRGRDFGGLYKQGTVEVADOLRGVWL 593
QY 586 ANIPWIDADHIGIWGWSFGGYLTSSKLEKDSGAPTLGVIITAPVSDMRFYDSMYTERMYKT 645
Db 594 KQPPWVDPARIGVQWNGSGGWTMLLAKASDSYACGVAGAPVTDWGLYDSHYTERYMDL 653
QY 646 LSTNEEGYETSAVRKTGDKNVGEG---GFLIQHGTGDDNVHFQNSAALVDLLMGDGVSP 702
Db 654 PARNDAGYREARV-----LTHIEGLRSPLLLIHGMADDNVLPFTNSTLSMALQKRG-QPF 707
QY 703 KLHSQWFTDSHGHS 717
Db 708 ELMT--YFGAKHGLS 720

RESULT 14
T37700
probable dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe) (fragment
C:Species: Schizosaccharomyces pombe
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T37700
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21737
A:Accession: T37700
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-743 <DEV>
A:Cross-references: UNIPARC:UPI000006A67C; EMBL:Z98596; PIDN:CAB11208.1; GSPDB:GN000066;
A:Experimental source: strain 972h-; cosmid c14C4
C:Genetics:
A:Gene: SPDB:SPAC14C4.15C
A:Map position: 1

Query Match 14.5%; Score 584; DB 2; Length 743;
Best Local Similarity 23.2%; Pred. No. 5.8e-30;
Matches 177; Conservative 154; Mismatches 287; Indels 144; Gaps 26;
QY 20 FNETVVKQAITPSTRSVQWLSGAEDGSYVY-----AAEDGSLTIENIVTNE 65
Db 6 FGESI-----FLPYHQDIEWITSTE-GTVLYYDQSTFSLFYPDPGKEYGS-NVDSLITS- 58
QY 66 SRTLIPADKIPITGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLSKSVPLM 125
Db 59 -----FVLTKNLHRKRYSSDMEYIAFSCSKRRWEHSYVEDVYLVERATGRIEHLA 110
QY 126 PQEGDIQYQAQSPVGNITAFVRENDLYWDNGT--VTRITDDGGPD-MFHGVDPDIYEE 182
Db 111 SDQSKKIIVVAESPIGHKLVYGLGSLNLFWESFSEPPVCITDQSLDGLFNGNSDWVYEE 170
QY 183 EILGDRYALWSPDPGEYLAYSFNETGVPTYTVQYMDNQETAPAYPWEKIRYPKVSQT 242
Db 171 EILQSSKAVWMSPDGNCGLSYLSIDDSKVPVHVLPEQLDSDKVEDQNRVNNFFHYSTPKDP 230
```

QY	243	NPTVTLISLNIASKEVKQAPIDA---FESTDLIIGEVAWJTDTHTTVAAKAFNRV--QDQ	297
Db	231	IPFVKL-FVNCFTDSGSEIEVDSSFPPLSTQHRVITDVAMAGNEKL-----MFVEVLRGNY	284
QY	298	QKVAVDTASNKATVISDRDGTG--WLDNLLSMKY-----IGPIKPSDKDAYIIDSDH	350
Db	285	ERVTSFLDLSSRKTIENTEVSHPALATSSHLKYLSPESLGNLKERVYRQYFL-----	339
QY	351	SGWAHLVFPVSGGERPLTKGDWEVTSIISIDQERQLVYVLTQHHSTERHLYSYST	410
Db	340	SNKKRIAIYELDNPVPIYLPVNLISFLSDLYL--INNTLYFTAISGSPFSRVRLCTKS	397
QY	411	FAVTPLVDDTVAAWYWSAFSANGSYIYLTTCGDPVYQELYYTNTSTKPLTIIDNAK---	467
Db	398	LILSEINIQISGLFGIKVNDQNYLLVNLVYGPPEIKQFIYSIHEDK-VSTNSDHSKNL	456
QY	468	-----VLBOIK-----DVALPNITYFELPLPSGETLNVQRLPPG	502
Db	457	PSDSSSTSLGKVKLELCSLETNEELIITKEFAFPSPV-PFKVIKVKNIITAYIQEIRPPN	515
QY	503	FSPDKKYPILFTPPYGGPCAQEVTKRWQ-----ALNFKA-----YVAS	539
Db	516	FNPRKRYPTFPHYGAPQSAVLTQKYEINDINELMASVYNFLVLIKVDIVDISVGGQHLFS	575
QY	540	DS-ELEVVTVDNRGTGFGKGRKFRSAVTRQLGLEAEQIYAAQQAANI--PWIDADHIGI	598
Db	576	DSHELIIKSW-----LELLRSYVDTPYIDRHRVGI	605
QY	599	WGSFGGYLTSKLVLEKDSGAPTLGVIITAPVSDMRFYDSMYTERYMKTLSTNEEG-YETSA	657
Db	606	WGSFGGYLTKILE-NADFIKTAQVAVPVTDMRYDAYISENLGAYSKQTTAIYDKTA	664
QY	658	VRKTDGKFNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGV	699
Db	665	VHYSENPRKL-CGLLVHLGTSDDNVHIENTMQLTKAMVEKGV	705
RESULT 15			
T25173			
hypoetical protein T23F1.7a - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T25173			
R:Wilkinson, J.			
submitted to the EMBL Data Library, October 1996			
A:Reference number: Z19990			
A:Accession: T25173			
A>Status: preliminary; translated from GB/EMBL/DBDJ			
A:Molecule type: DNA			
A:Residues: 1-779 <WIL>			
A:Cross-references: UNIPROT:O18119; UNIPARC:UPI000002A220; EMBL:Z81129; PIDN:CAB03411.1;			
A:Experimental source: clone T23F1			
C:Genetics:			
A:Gene: CESP:T23F1.7a			
A:Map position: 5			
A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1			
C:Superfamily: dipeptidyl-peptidase IV			
Query Match 13.5%; Score 545; DB 2; Length 779;			
Best Local Similarity 28.0%; Pred. No. 2.1e-27;			
Matches 180; Conservative 105; Mismatches 245; Indels 114; Gaps 27;			
QY	128	QEGDIQYQAWSVPGNTIAFVRENDLYVW---DNGVTTRITDDGGPDMFHVDPDWIYEEI	184
Db	161	EESVIQAQKNGKFNDFVFSNEPKIYQSSPEEGLTRVS-NGGHEVTVDGLFDWIYEEI	219
QY	185	LGDRVALWFSPPGEVLAYLSNETGVPTVTVOYMDNQELADAPYEWELKIRYKPV-SQTN	243
Db	220	FGRKDAMWWSSTKQDLAVASYDNHILTKVNSLKY---HRLP-YPIDTNFHYPKTFKVL	275
QY	244	PTVTLISLNIASKEVKQAPIDAFESTDL-IIGEVAWLTDTHTTVAAKAF-NRVQDQKVV	301
Db	276	PTVTLISWNKTEQSRQLDVQLKSLSVHYLLAVKMLEINGTEQLVSVMTNRNYQNEVALT	335

QY	302	AVD-----TASNKATVISDRDGTGWLNDNLLSMKYIGPIKPSDKDAYIIDIS	348
Db	336	ICDWDTAICRLEFEYKYASKRWTHDDPHSITSFEDTLFF-----LLPHDK-----	381
QY	349	DHSGWAHLVFPVSGGE---PIPLTKGDWEVTSIISIDQERQLVYVLTQHHSTERHLYS	405
Db	382	RNAPFOOQVASLRUSHQQLRTPKFLNLGCEYDVDTISINGINKETRTIPFAAAPKPSHSLFS	441
QY	406	VSYSTPAVTPLVDDTVAAWYWSAFSANGY-----YILTYGGP	443
Db	442	YS-----LADE-----SNSAYCISCSIKNCTWAOAQMDDQMKTAVISCKGP	483
QY	444	DVPYQ---ELYTNTSTKPLRTITDNAKVL-----EQIKDYALPNITYFELPLPSGETLN	494
Db	484	AAPHTAIVNLTRMDSDK-----TEHANLLYDKTYQNRVEEAGLPVITKETIKISDDFDAL	539
QY	495	VMQRLPPG-FSPDK--KYPILFTPPYGGPCAQEVTKRWQALNFKAYYVASDSELEYVVTWVD	551
Db	540	IKLSIPKDIYNRDKHQAIPLIVHVYGGPNDQN-TKEATQIGIEEVVASASQAAIL--RID	596
QY	552	NRGTFGKGRKFRSAVTRQLGLEAEQIYAAQQAANI--PWIDADHIGIWMGSGGYLT-	608
Db	597	GRSGGGRGWKYRSIIYQGLTVEVEQIKAIKVVLRLYRHLLDARRVAVFGWSYGGFMTL	656
QY	609	SKYLEKDSGAPTLGVITAPVSDMRFYDSMYTERYMKTLSTNEEGYETSAYRKTDGKQVE	668
Db	657	SMVNEAPEQPFKCAVSVAPVTNFAYTDATYTERYMGDAPL--ESY-SDVTKKLDNFKSTR	713
QY	669	GGFLIOHGTGDDNVHFQNSAALVDLLMGDGVS-----PEKLHS	706
Db	714	--LLLHMLGLDDNVHFQNSAILIDELQNRGVDFDLVMVYFNOAHS	755

Search completed: January 9, 2006, 03:20:54  
Job time : 27 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 00:27:19 ; Search time 85 Seconds  
(without alignments)  
3902.717 Million cell updates/sec

Title: US-10-723-807-2\_COPY\_17\_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTEGSKRLTF.....EKNRQTVLMHQWTKKDLLE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 21:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4033	100.0	771	2	Aaw89589 Aspergill
2	3886	96.4	771	2	Aaw97798 Aspergill
3	2914	72.3	764	8	AdS23769 Bacterial
4	2610.5	64.7	775	9	AdY51819 T. rubrum
5	1269	31.5	790	8	AdN21150 Bacterial
6	1162	28.8	793	8	AdN19590 Bacterial
7	1118	27.7	634	8	AdS24271 Bacterial
8	1054	26.1	818	8	AdS43599 Bacterial
9	1020	25.3	504	5	AdI17327 Polypept
10	987.5	24.5	759	2	Aar54612 Delta3-9
11	987.5	24.5	766	2	AAR40909 Sequence
12	987.5	24.5	766	5	ABB08991 Human dip
13	987.5	24.5	766	5	AG78417 Human dip
14	987.5	24.5	766	7	ADD27855 Human dip
15	987.5	24.5	766	7	ADD46934 Human Pro
16	987.5	24.5	766	7	ADN39272 Cancer/an
17	987.5	24.5	766	8	ADJ83981 Human ful
18	987.5	24.5	766	8	ADJ75313 Marker ge
19	987.5	24.5	766	8	ADQ19398 Human PRO
20	987.5	24.5	766	8	ADO19806 Human PRO
21	987.5	24.5	766	8	ADO71612 Amino aci
22	987.5	24.5	766	8	ADO71644 Amino aci
23	987.5	24.5	766	8	ABM80355 Tumour-as
24	987.5	24.5	766	8	ADP54458 Human PRO

25	987.5	24.5	766	8	ADU06688	Novel bro
26	987.5	24.5	766	8	ADV25525	Human dip
27	987.5	24.5	766	9	ADY15161	PRO polyp
28	987.5	24.5	766	9	ADY16580	PRO polyp
29	987.5	24.5	766	9	ADZ14038	Human dip
30	987.5	24.5	766	9	AEb94223	CD26/dipe
31	986.5	24.5	766	2	AAR54611	Native CD
32	982	24.3	734	9	AEb94218	Human sol
33	982	24.3	760	7	ADN95552	Human BEC
34	982	24.3	760	8	ADQ21351	Human sof
35	982	24.3	760	9	AEb94159	Human wil
36	981.5	24.3	723	9	AEb94227	Human sol
37	981.5	24.3	736	8	ADO40240	Human DPP
38	981.5	24.3	766	5	ABG61910	Prostate
39	981.5	24.3	766	5	AAO15555	Human dip
40	981.5	24.3	766	6	ABP55629	Human dpp
41	981.5	24.3	766	6	ABP56700	Human liv
42	981.5	24.3	766	7	ADD14045	Human src
43	981.5	24.3	766	7	ADN39604	Cancer/an
44	981.5	24.3	766	8	ADO19400	Human PRO
45	981.5	24.3	766	8	ADQ80365	Dipeptidy

## ALIGNMENTS

## RESULT 1

Aaw89589

ID Aaw89589 standard; protein; 771 AA.

XX Aaw89589;

DT 17-MAR-1999 (first entry)

DE Aspergillus oryzae dipeptidyl aminopeptidase #1.

XX Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;

KW flavour enhancer; palatability; mouthfeel; aroma; crust colour; baking;

KW animal feed additive; hydrolysis.

XX Aspergillus oryzae.

PN WO9851803-Al.

XX 19-NOV-1998.

PF 12-MAY-1998; 98WO-US009629.

PR 16-MAY-1997; 97US-00857884.

PR 20-OCT-1997; 97US-0062892P.

(NOVO ) NOVO NORDISK BIOTECH INC.

Blinkovsky A, Brown K, Rey MW, Klotz A, Byun T;

WPI; 1999-045232/04.

N-PSDB; AAX00013.

XX New dipeptidyl aminopeptidase from Aspergillus oryzae - used to produce

PT protein hydrolysates enriched in particular amino acids, useful as

PT flavour enhancers, e.g. in doughs.

XX Claim 1; Page 56-57; 77pp; English.

XX The present sequence represents dipeptidyl aminopeptidase (DPAP) from

CC Aspergillus oryzae. DPAP acts synergistically with an aminopeptidase (AP)

CC to hydrolyse polypeptides, producing protein hydrolysate (pH), useful in

CC foods as flavour enhancer, e.g. in baked goods, enriched in: (a) Ala,

CC Arg Asp, Gly and/or Val, or (b), if the substrate has been deamidated,

CC in Glu (free and/or peptide bound), in which case products are useful as

CC animal feed additives. DPAP can also be used in flavour- improving

CC compositions (optionally containing AP) and in dough pre-mixes, also for

CC deactivating enzymes and for converting precursors to mature proteins.

CC DPAP increases the level of hydrolysis of proteins and thus of flavour  
CC development, and a mixture with AP may hydrolyse tripeptides that are  
CC resistant to either enzyme used alone. PH have improved solubility,  
CC emulsifying and foaming properties, and products containing them have  
CC better flavour, palatability, mouthfeel, aroma and crust colour  
XX  
SQ Sequence 771 AA;

Query Match 100.0%; Score 4033; DB 2; Length 771;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LDVPRKPHAPTGEKSKLTFTNETVVKQAITPTSSVQWLSGAEDGSVYAAEDGSLTIEN 60  
Db 17 LDVPRKPHAPTGEKSKLTFTNETVVKQAITPTSSVQWLSGAEDGSVYAAEDGSLTIEN 76  
Qy 61 IVTNESRTLIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLK 120  
Db 77 IVTNESRTLIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLK 136  
Qy 121 SVPLMPDQEGDIQYQAQSPVGNITAFVRENDLYVMDNGTVTRITDDGGPDMFHGVPDWIY 180  
Db 137 SVPLMPDQEGDIQYQAQSPVGNITAFVRENDLYVMDNGTVTRITDDGGPDMFHGVPDWIY 196  
Qy 181 EEEILGDRYALWFSPDGEYLAYSFNETGVPYTVQYMDNQEIAPAYPWEKIRYPKVS 240  
Db 197 EEEILGDRYALWFSPDGEYLAYSFNETGVPYTVQYMDNQEIAPAYPWEKIRYPKVS 256  
Qy 241 QTNPTVTLSLNIASKEVKQAPIDAFESTDLICEVAWLTDTHTTVAAKAFNRVQDOQKV 300  
Db 257 QTNPTVTLSLNIASKEVKQAPIDAFESTDLICEVAWLTDTHTTVAAKAFNRVQDOQKV 316  
Qy 301 VAVDTASNKATVISDRDGTGWLNLNLSMKYIGIPKSPDKDAYYIDISDHSGWAHLFLFP 360  
Db 317 VAVDTASNKATVISDRDGTGWLNLNLSMKYIGIPKSPDKDAYYIDISDHSGWAHLFLFP 376  
Qy 361 VSGGEPIPLTKGDWEVTSILSIDERQLVYLSLTHQHSSTERHLYSVSYSTFAVTPLVDDT 420  
Db 377 VSGGEPIPLTKGDWEVTSILSIDERQLVYLSLTHQHSSTERHLYSVSYSTFAVTPLVDDT 436  
Qy 421 VAAWKSASFANSQVYIITLYGGDPVPYQELVTTNSTKPLRTITDNAKVLEQIKDYALPNI 480  
Db 437 VAAWKSASFANSQVYIITLYGGDPVPYQELVTTNSTKPLRTITDNAKVLEQIKDYALPNI 496  
Qy 481 TYFELPLPSGETLNMORLPPGFPDKKYPILFTFYGGPGAQEVTKRQALNFKAYVASD 540  
Db 497 TYFELPLPSGETLNMORLPPGFPDKKYPILFTFYGGPGAQEVTKRQALNFKAYVASD 556  
Qy 541 SELBYVTWVNRGTGPKGRFSAVTRQLGLLEAEQIYAAQQAANIPIWDADHIGIWG 600  
Db 557 SELBYVTWVNRGTGPKGRFSAVTRQLGLLEAEQIYAAQQAANIPIWDADHIGIWG 616  
Qy 601 WSPFGYLTSKVLEKDSGAFTLGVTAPVSDWRFRVDSMYTERYMKTLSTNEEGYTSAVRK 660  
Db 617 WSPFGYLTSKVLEKDSGAFTLGVTAPVSDWRFRVDSMYTERYMKTLSTNEEGYTSAVRK 676  
Qy 661 TDGFKNVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGSIYHG 720  
Db 677 TDGFKNVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGSIYHG 736  
Qy 721 GGVELYKQARKLYOEKNRQTVQLMHWTKKDLLEE 755  
Db 737 GGVELYKQARKLYOEKNRQTVQLMHWTKKDLLEE 771

## RESULT 2

AAW97798

ID AAW97798 standard; protein; 771 AA.

XX

AC AAW97798;

XX

DT 21-MAY-1999 (first entry)

XX

DE Aspergillus oryzae prolyl-dipeptidyl-peptidase DPPIV.  
XX Prolyl-dipeptidyl-peptidase; DPPIV; protein hydrolysate; food.  
XX Aspergillus oryzae.  
XX  
PH Key Location/Qualifiers  
FT Peptide 1..16  
FT Protein /note= "signal peptide, specifically claimed in Claim 4"  
FT /note= "mature protein, specifically claimed in Claim 1"  
XX  
XX WO9902705-A1.  
XX 21-JAN-1999.  
XX 06-MAY-1998; 98WO-EP002799.  
XX 05-JUL-1997; 97EP-00111377.  
XX (NEST ) SOC PROD NESTLE SA.  
XX Monod M, Doumas A, Affolter M, Van Den Broek P;  
XX WPI; 1999-120906/10.  
XX N-PSDB; AAX07327.  
XX  
XX New Aspergillus oryzae prolyl-dipeptidyl-peptidase enzyme (DPPIV) -  
XX having a high level of hydrolysing specificity towards proteins starting  
XX with X-Pro, where X is any amino acid.  
XX  
XX Claim 1; Page 27-29; 41pp; English.  
XX  
XX This is the amino acid sequence of the prolyl-dipeptidyl-peptidase  
XX (DPPIV) of Aspergillus niger strain 44. DPPIV is a secreted enzyme having  
XX a high level of hydrolysing specificity towards proteins starting with X-  
XX Pro, where X is any amino acid. The dppiv gene (see AAX07327) was  
XX isolated by screening a genomic DNA library with a DNA fragment containing  
XX the dppiv gene of Aspergillus fumigatus. The invention also provides host  
XX cells able to over-express the enzyme, providing at least 50 mu of DPPIV  
XX activity/ml of supernatant when grown in minimal medium containing 1  
XX wt./vol% of wheat gluten, as well as Aspergillus cells having integrated  
XX multiple copies of the dppiv gene promoter, recombinant prolyl-dipeptidyl  
XX -peptidase having the leader sequence of DPPIV, and a method for  
XX producing the enzyme by cultivating recombinant cells. The DPPIV enzyme  
XX or cells expressing it can be used, optionally in conjunction with  
XX another proliase, to hydrolyse protein containing materials. Food  
XX products can be a milk substitute for infants, a hydrolysed vegetable  
XX protein (koji), a leguminous or oleaginous plant, or cooked or roasted  
XX cereal sources. If the DPPIV activity is combined with other proteolytic  
XX activities, a high degree of hydrolysis may be obtained leading to a non-  
XX bitter flavour and a significantly lower allergenicity than unhydrolysed  
XX proteins  
XX  
SQ Sequence 771 AA;

Query Match 96.4%; Score 3886; DB 2; Length 771;  
Best Local Similarity 96.8%; Pred. No. 6.1e-311;  
Matches 733; Conservative 1; Mismatches 19; Indels 4; Gaps 2;

Qy 1 LDVPRKPHAPTGEKSKLTFTNETVVKQAITPTSSVQWLSGAEDGSVYAAEDGSLTIEN 60  
Db 17 LDVPRKPHAPTGEKSKLTFTNETVVKQAITPTSSVQWLSGAEDGS--LRVGRGRROSHH 74  
Qy 61 IVTNESRT--LIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVES 118  
Db 75 REHHQRVTHAHFADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVES 134  
Qy 119 LKSVPLMPDQEGDIQYQAQSPVGNITAFVRENDLYVMDNGTVTRITDDGGPDMFHGVPDW 178  
Db 135 LKSVPLMPDQEGDIQYQAQSPVGNITAFVRENDLYVMDNGTVTRITDDGGPDMFHGVPDW 194  
Qy 179 IYBEEILGDRYALWFSPDGEYLAYSFNETGVPYTVQYMDNQEIAPAYPWEKIRYPK 238



```
Db 195 IVEBEILGDRVALWSPDGEYLAISFNETGVPTTVQYMDNQEIADAPWELKIRYPK 254
Qy 239 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 298
Db 255 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 314
Qy 299 KVAVADTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSKDQAYIIDISDHSGWAHLYL 358
Db 315 KVAVADTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSKDQAYIIDISDHSGWAHLYL 374
Qy 359 FVPSGGEPIPLTKGDWEVTSILSDOERQLVYLSSTQHSTHERHLYSVSYSTFAVTPDVD 418
Db 375 FVPSGGEPIPLTKGDWEVTSILSDOERQLVYLSSTQHSTHERHLYSVSYSTFAVTPDVD 434
Qy 419 DTVAAYWASFSANSFYILTYGGPDVPYQELIYTTNSTKPLRTITDIAKVLQIKDYALP 478
Db 435 DTVAAYWASFSANSFYILTYGGPDVPYQELIYTTNSTKPLRTITDIAKVLQIKDYALP 494
Qy 479 NITYFELPLPSGETLNVMQRLPPGFSQDKYKYPILFTPYGGCAQVTKRQALNFKAYVA 538
Db 495 NITYFELPLPSGETLNVMQRLPPGFSQDKYKYPILFTPYGGCAQVTKRQALNFKAYVA 554
Qy 539 SDSELEYVTWVDNRGTGPKGRKFSAVTRQLGLEAEQIYAAQQAANIPWIDADHIGI 598
Db 555 SDSELEYVTWVDNRGTGPKGRKFSAVTRQLGLEAEQIYAAQQAANIPWIDADHIGI 614
Qy 599 WQWSEGGYLTSTKLEKSGAFTLGIVITAPVSDWRFYDSMYTRYMYKTLSTNEEGYETSAV 658
Db 615 WQWSEGGYLTSTKLEKSGAFTLGIVITAPVSDWRFYDSMYTRYMYKTLSTNEEGYETSAV 674
Qy 659 RTIDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSDHGISY 718
Db 675 RTIDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSDHGISY 734
Qy 719 HGGGVFLYKQLARKLYQEKNRQTQVLMHQWTKKDL EE 755
Db 735 HGGGVFLYKQLARKLYQEKNRQTQVLMHQWTKKDL EE 771

RESULT 3
ID ADS23769 standard; protein; 764 AA.
XX ADS23769;
AC ADS23769;
DT 02-DEC-2004 (first entry)
XX Bacterial polypeptide #12802.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX Bacteria.
XX OS
XX PN
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
```

```
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 12802; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or by
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 764 AA;
Query Match 72.3%; Score 2914; DB 8; Length 764;
Best Local Similarity 69.3%; Pred. No. 9.2e-231;
Matches 524; Conservative 103; Mismatches 113; Indels 16; Gaps 4;
Qy 4 PRKHAPTEGEGSKRLTFNETVVKQAITPTSRVQVLSGAEQSGSYV-----YAAEDGSLTI 58
Db 17 PWRPREPRAAGSKRLTFNETVISAALSPPSSISVQVIATENDGGCLCVPGGKWHQDXEHCH 76
Qy 59 ENIVNESRTLLPADKIPGKEAFNVTWHPDLSVLSWASNHTKQYRHSFFADYVYQDVES 118
Db 77 QPIADNRSAEKIPAD-----AYSTWISPDLSAVLMATNYTKQYRHSFFADYVYQDVET 129
Qy 119 LKSVPLMPDQEGDIQYAOQWSPVGNITAFVRENDLYVNDGTVTTRITDDGPGDMFHGVPDW 178
Db 130 LETVPLVEDMVGDIQYAEWSPSGDSIAFVRGNLMTWSDGTVTALTDCGPGDMFHGVPDW 189
Qy 179 IVEBEILGDRVALWSPDGEYLAISFNETGVPTTVQYMDNQEIADAPWELKIRYPK 238
Db 190 IVEBEILGDRVALWSPDSELLAFLTFNETGVPTTVQYMDNQEIADAPWELKIRYPK 249
Qy 239 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 298
Db 250 VSETNPTVKLNILQJLSDNTVTSTIPIDVDFPSELIVGEVAWLTDTHTTVAAKAFNRVQDES 309
Qy 299 KVAVADTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSDKD---AYYIDISDHSGWAH 355
Db 310 KVVIVETASGETKIAHERDGTGMDLNLNLSISVGFALGSGDASSAYYVLDSDHSGWTH 369
Qy 356 LYLFPVSGGEPIPLTKGDWEVTSILSDOERQLVYLSSTQHSTHERHLYSVSYSTFAVTP 415
Db 370 LYLFTSTGGDPIPLTEGEWEVTSIVSDOERELFYLLSTQHHSTHERHLYSVSYRTFEITP 429
Qy 416 LVDDTVAAYWASFSANSFYILTYGGPDVPYQELIYTTNSTKPLRTITDIAKVLQIKDY 475
Db 430 LVDDTVEAYWSVFSAKAGYIILTYAGPSVPYQELIYTVNQAPLRTLTSTNAALIEKLEEY 489
```

476	Qy	ALPNITYPELPLP	SGSETLNV	MQRLP	DPGSP	DPK	KYB	ILFT	PYG	GGP	GAQ	EVTK	QWQ	ALN	FKA	535
490	Db	ALPNISYFELE	IPSGEKUN	VMQRL	VPVGS	PDPK	KYF	PVLFT	PYG	GGP	GAQ	EVSK	QWQ	SLD	NFA	549
536	Qy	YVASDSELEY	TWTV	VDNR	GTG	FKGR	KFS	SAV	TROL	GLLE	EAED	POI	YAA	QQA	ANI	595
550	Db	YIASDPELEY	TWTV	VDNR	GTG	GYR	GREF	RS	LVA	QK	GLKE	AEED	OVV	AAQ	AKL	609
596	Qy	IGIMGWSFG	GYLTS	KVL	EKDS	GAFT	LGV	IT	APV	SD	WR	FVDS	MYT	ERY	MKT	655
610	Db	IAIMGWSYGG	YLT	TKVLE	TD	SGAF	SLGL	IT	APV	SD	WR	LYDS	MYT	ERY	MKT	669
656	Qy	SAVRKTGCF	KNV	EGFL	IOH	GTG	DDNV	HF	QNS	AA	LVD	ILMG	DGV	SG	PEK	715
670	Db	TAIRHTDGF	KNV	EGFL	IOH	GTG	DDNV	HF	QNS	AA	LVD	ILMG	DGV	SG	PEK	729
716	Qy	ISYHGGGV	FLYK	Q	LARK	LYQ	EKN	RQ	TQV	LMH	QW	TKK				751
730	Db	IRYNGGNV	FLYR	Q	LAOR	LYQ	EKN	RK	KE	-	Q	HOWS	SR			764

RESULT 4

RESCUE 1  
ADY51819

ID ADY51819



AC ADY51819;

XX  
DT 19-MAY-2005 (first entry)

XX

DE T. rubrum dipept

XX Fungicide: Tr

KW Fungicide; Immunosuppressive;  
KW Metabolic; Protease Inhibitor

KW  
immune disorder; fungal infection

KW dipeptidylpeptidase

XX

OS *Trichophyton rubrum*.

XX	1
XX	2
XX	3
XX	4
XX	5
XX	6
XX	7
XX	8
XX	9
XX	10
XX	11
XX	12
XX	13
XX	14
XX	15
XX	16
XX	17
XX	18
XX	19
XX	20
XX	21
XX	22
XX	23
XX	24
XX	25
XX	26
XX	27
XX	28
XX	29
XX	30
XX	31
XX	32
XX	33
XX	34
XX	35
XX	36
XX	37
XX	38
XX	39
XX	40
XX	41
XX	42
XX	43
XX	44
XX	45
XX	46
XX	47
XX	48
XX	49
XX	50
XX	51
XX	52
XX	53
XX	54
XX	55
XX	56
XX	57
XX	58
XX	59
XX	60
XX	61
XX	62
XX	63
XX	64
XX	65
XX	66
XX	67
XX	68
XX	69
XX	70
XX	71
XX	72
XX	73
XX	74
XX	75
XX	76
XX	77
XX	78
XX	79
XX	80
XX	81
XX	82
XX	83
XX	84
XX	85
XX	86
XX	87
XX	88
XX	89
XX	90
XX	91
XX	92
XX	93
XX	94
XX	95
XX	96
XX	97
XX	98
XX	99
XX	100

PN WO2005019251-A2.  
yy

03-MAR-2005

XX  
CTC-9704

PF 25-AUG-2004; 2004WO

XX

PR 25-AUG-2003; 2003US-0498318P.

XX  
PA (ATHE-) ATHERIS LAB.

PA (MONO/) MONOD M.

Category	Item	Value
A. General Information	1. Name	John Doe
	2. Address	123 Main St, New York, NY 10001
	3. Phone	(212) 555-1234
	4. Email	john.doe@example.com
	5. Date of Birth	01/01/1980
	6. Gender	Male
	7. Marital Status	Single
	8. Education	High School Graduate
	9. Occupation	Software Engineer
	10. Current Employer	ABC Corporation
B. Financial Information	11. Annual Income	\$75,000
	12. Monthly Expenses	\$2,500
	13. Savings	\$10,000
	14. Debt	\$5,000
	15. Credit Score	720
	16. Assets	Car, House
	17. Liabilities	Mortgage, Student Loan
	18. Net Worth	\$15,000
	19. Investment	Stocks, Bonds
	20. Insurance	Life, Health, Auto
C. Personal Information	21. Hobbies	Reading, Hiking
	22. Pets	1 Dog, 2 Cats
	23. Travel	Europe, Asia
	24. Languages	English, Spanish
	25. Skills	Programming, Writing
	26. Interests	Technology, Sports
	27. Values	Family, Integrity
	28. Goals	Home Ownership, Retirement
	29. Challenges	Job Market, Economy
	30. Summary	Overall Financial Health: Good

PI  
YY  
STOCKLIN R, Grouzmann E;

DR WPT: 2005-202598/21.

DR N-PSDB; ADY51818.

XX

PT New polypeptides and encoding nucleic acid molecules useful for treatment

PT diseases (e.g. fungal infections, celiac disease, digestive tract

malabsorption, s

Query Match	64.7%	Score	2610.5	DB	9	Length	775
Best Local Similarity	61.6%	Pred.	No. 1e-205				
Matches	463	Conservative	134	Mismatches	148	Indels	7
Gaps	4						
QY	4	PRKPHAPTGGSKRLTFNEVTKVKAATPTPSRSVOWLSGAEDGSYVYAAEDGSLTIENIVT	63				
DB	19	PRSPRSTGGGNKLLIYKECVPRATISPRSTSLAWINSEEDGRYISQSDGAILQNIYT	78				
QY	64	NESTRITPADKIPITGKEAFNWIYHPDLSSYLWASNHTKQYRHSFFADYYVQDVESLKSV	123				
DB	79	NTNKTLVAADKVPKG--YDYWFKPDLSAVLMATNVTQYRHSYFANYFILDIIKKSGLTP	136				
QY	124	LMFQDGDIOYAWQSPVGNITIAVRENDLYVWDNGVTVITITDDGGDPMFGHPDWTIYEE	183				
DB	137	LAQDAQDGIQYAWQSPMNINSIAVVRXNDLIYNNNGTKRITENGGPDIIFNGVPDWTYEE	196				
QY	184	ILGDRVALWFSPPGEYLAIVLSFNETGVPTVTVQYVMDNOBIAPAYPWLKIRPKVSOQN	243				
DB	197	IFGDRFALWFSPPGEYLAIVLRNETGVPIYTPYKXKQKIAPAYPRELEIRPKVSAKN	256				
QY	244	PTVTLSLINTASKEVQKQAPIDAFESTDLIIGEVAMLTDTHTTVAAKAFNRVQQQKVAV	303				
DB	257	PTVQFHLNTASSQETTIPVTAPPENDLVIGEVAWLSGGHDSVAYRAFNRVDREKIVSV	316				
QY	304	DTASNKATVLSRDGDTGMDNLNLSMKYGPICKPSDKDAYIIDISDHSGWAHLIYLPVSG	363				
DB	317	KVESKESKVRERDGTGDTGWINLLNLSYISGNVNGKE---YYVDISDASGWAHLIYLPVDG	373				
QY	364	GEPIPTKGDEWETSILSIDOERQLVYVLSSTQHSHTERHLYSVSYSTFAVTPLVDDTVAA	423				
DB	374	GKEALTKEWEVVALIKVDTKKKLIYFTSKVHSTTRHVSVSYDTKVMTPLVNDKEAA	433				
QY	424	YWGAASFANSYYILTYGGDPVYQBELYTT-NSTKPLRTITDNKVLQIKYVALPNITY	482				
DB	434	YVTASFSAKGGYILSYQGNPNVPYQELYSKDSKSKPLKTTITSNDALLEKLEYKLPKVSF	493				
QY	483	FELPLPSGETLNVWQRLPGFSPDKPKYPIILFTPYGGGQAEVTKRWAALNFKAYVASDSE	542				
DB	494	FEIKLPSGETLNVKQRLPPNFNPHKKYPVLVFTPYGGGQAEVVSQAWNSLDFKSYITSDPE	553				
QY	543	LEYVTWTDVNRGTGFKGRKFRSATVTRQLGLEAEQIYAAQQAANTFWIDADHIGWGWS	602				
DB	554	LEYVTWTDVNRGTGFKGRKFRSAVAKRLGFLEAQDVFAAKEVLKMRWADKOHIGIWGXS	613				
QY	603	FGVYLTSKVLKESGAPTLGVIITAPVSDMRFPYDSMTTERYMKTLSTNVEGYESAVRKT	662				
DB	614	YGGFLTAKTLETDSGVFTFGISTAPVSDFRLYDSMYTERYMKTVELNADGYSETAVHKVD	673				
QY	663	GFKNVEGGFLIOHGTGDDNVHFONSAAALVDLLMGDGVSPKHLHSQMFSTDHGISYVHGGG	722				
DB	674	GFKNLKGHYLIIHGTGDDNVHFOAVALVSLNTLMNGGVTTADKLTQTQFTSDHGIRDMDS	733				
QY	723	VFLYQLARKLYQEKNRQTO-VLMHQWTKDGL	753				
DB	734	TYQYKQLSKMVIYDQQRPRPSPPHQWSKRVL	765				

RESULT 5

ADN21150

ID ADN21150 standard; protein; 790 AA.

ADN21150;  
ACXX  
XX  
—

DT 02-DEC-2004 (first entry)

XX Recombinant DNA construct; transformed plant; improved plant property;  
DE cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
XX pathogen tolerance; pest tolerance; plant disease resistance;  
XX cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW Bacterial polypeptide #3803.

nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
bacterial polypeptide.  
Bacteria.  
US2003233675-A1.  
18-DEC-2003.  
20-FEB-2003; 2003US-00369493.  
21-FEB-2002; 2002US-0360039P.  
(CAOY/) CAO Y.  
(HINK/) HINKLE G J.  
(SLAT/) SLATER S C.  
(CHEN/) CHEN X.  
(GOLD/) GOLDMAN B S.  
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
WPI; 2004-061375/06.  
New recombinant DNA construct comprising a promoter positioned to provide  
for expression of a polynucleotide encoding a polypeptide from a  
microbial source, useful for producing plants with improved properties.  
Claim 1; SEQ ID NO 3803; 122pp; English.  
The invention relates to a recombinant DNA construct comprising a  
promoter functional in a plant cell, where the promoter is positioned to  
provide for expression of a polynucleotide encoding a polypeptide from a  
microbial source. The invention also relates to a transformed plant  
comprising the recombinant DNA construct and a method of producing a  
transformed plant having an improved property. The plant is a crop plant  
such as maize or soybean. The method of producing a transformed plant  
having an improved property comprises transforming a plant with the  
recombinant DNA construct and growing the transformed plant, where the  
polynucleotide or polypeptide is useful for improving plant properties.  
The recombinant DNA construct is useful for producing plants with  
improved plant properties, e.g. improved cold, heat or drought tolerance,  
tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
increased resistance to plant disease, better growth rate by modification  
of the cell cycle pathway with plant growth regulators, increased rate of  
homologous recombination, modified seed oil or protein yield and/or  
content, improved yield by modification of carbohydrate, nitrogen or  
phosphorus use and/or uptake, by modification of photosynthesis or by  
providing improved plant growth and development under at least one stress  
condition, improved lignin production or improved galactomannan  
production. This sequence represents a bacterial polypeptide used in the  
scope of the invention. Note: The sequence data for this patent did not  
form part of the printed specification but was obtained in electronic  
format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
Sequence 790 AA;  
Query Match 31.5%; Score 1269; DB 8; Length 790;  
Best Local Similarity 38.7%; Pred. No. 4.9e-95;  
Matches 296; Conservative 132; Mismatches 271; Indels 66; Gaps 29;  
QY 7 PHAPTGEKS-KRLTFNFTVVKQAITPTSRVQWLSG--AEDG-----SYVAAED 53  
DB 34 PQATVTHGSGKVTLDQ-VLNNWRKASHSISWAGNAGEDGLLLEKEGANKDYLVVEDV 92  
QY 54 GSLTNIENVNSTRILIPADKIPTGKEAFN--YWIHP-----DLSSVLWASNHTKQVR 104  
DB 93 RAQNPSSVEASKETLI-KDKL---FEFANKTYW--PTVTPSRDLKKVLLATDVQNNWR 146  
QY 105 HSPFADYVVDVSLKSVPLMP--DQEGDIQYAWQSPVGNVIAFVRENDLYV--WDNGTVT 161  
DB 147 HSYAVYVWIPDVETQQAEPVLPVADARLQLASWSPSTSDAIVYTRDNMFRLKLDSDKIV 206  
QY 162 RITDDGDPDFHGVDPDWIYEEBILGDRYALWFSFDGDEYLAISFNETGVFTYVQYWM-- 219

Db 207 QITRDGSADVFNVPDWMYEEBVLASGVATWMSDGNVAFURTNETGVPEYPIQYFVSR 266  
QY 220 -DNQEIAPA--YPWELKIRYPKVQSQTNTPTVTLNLNIAASKVKQAPIDA-FESTDLIIG 274  
Db 267 PSGEFPKEEENYPEVRQIKYKAGAHNPIDVLKFDYDKRGDVFSVDISGRADDDRLIT 326  
QY 275 EVANLTDTHTTAAKAFNRVQDQKQVAVDTASN-----KATVISDRDGTGWLONLLSM 329  
Db 327 EVIM--AGKQVLIKETNRVSDVMRVLVDVGSRTGKAVRTVDVNDIDG--GWFEISHKT 381  
QY 330 KYIGPIKPSD--KDAYYIDIDSHSWAHL-YLFPVSGGEPIPLTKGDMEVTSILS-IDOE 385  
Db 382 KFI-PADPANGRPDDGYVDTIHNNGDHLAYFTPLDNPFPIMLTSGDYEVVDAPSVDLQ 440  
QY 386 ROLVYVLTSTOHHSTERHLYSVSYTFAVTPLVDVTVAAWYWSASFANSYGYILTYGPDV 445  
Db 441 RNLVYFVSTKESSIORHYQVVKLTGEDMTPTVDTSKGYVYLSFSTGAGYALVSQGPNI 500  
QY 446 PYQELTYT--NSTKPLRTITDQNAKVLQIKQYALPNITYFELPLPSGETLNVMQRLPPGF 503  
Db 501 PMQKVISITPSNPKYEHVVEENKDLAAKAKHELPIYGTINV-DGVELNVIERRPPHP 559  
QY 504 SPDKKYPIILFTPYGPGAQEVTKRQALNFKAYVASDSLELYVTWVTDNRGTGFKGRKR 563  
Db 560 DNKKKYPVLFOQYSGFVSQTKTF-AVDFOQFVAAG--LGICYCVTDGRGTGFGIRKMR 616  
QY 564 SAVTRQLGLLEADQIYAAQQAANIPWIDADHIGTWGSGFGLYTSKVLKDSG-AFTLG 622  
Db 617 VTIRNGLGWESHDQIAAKGHWAKDYIDEDRLAIWGSYGGYMTLKTLEQDAGQTFKYG 676  
QY 623 VITAPVSDWRFFYDSMYTERYMKTLSTNBEGETSAVRKTDGP-KNVEGGFLIOHGTGDDN 691  
Db 677 MAVAPVTDWRFFYDSIYTERYMETPQTNPEGYESAAVTNVLTALSQNVR--FLLMHGVADDN 734  
QY 682 VHFQNSAALVDLLMGDGVSPKHLHSQWFTDSDHSGISYHGGGVFLY 726  
Db 735 VHMQNSLTLLDAL--DQRSVENYDVQVFPDSDHGIYFHANRIVF 777

RESULT 6  
ADN19590  
ID ADN19590 standard; protein; 793 AA.  
XX ADN19590;  
XX 02-DEC-2004 (first entry)  
XX Bacterial polypeptide #2243.  
XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polypeptide.  
XX Bacteria.  
XX US2003233675-A1.  
XX 18-DEC-2003.  
XX 20-FEB-2003; 2003US-00369493.  
XX 21-FEB-2002; 2002US-0360039P.  
XX (CAOY/) CAO Y.  
XX (HINK/) HINKLE G J.  
XX (SLAT/) SLATER S C.  
XX (CHEN/) CHEN X.  
XX (GOLD/) GOLDMAN B S.





Db 48 IPHSNTPDYQEPNSNYNDGKLKVSFS-VVRNNTFQPKYHELQWISDNKIESNDLGLYV 106  
Qy 49 YAAEDGSLTIENIVNESRTLIPADKIPTGKEAFNYWIH-----PDLSSVLW 95  
Db 107 TFMND-SYVVKSVYDDSNVSLLEGKT-----FIHNQNLTVESITASPDKLRLI 156  
Qy 96 ASNHTKQYRSHFFADYVYQDVESIKSVPLMPDQGDQIYAQOWSPVGNVTIAFVRENDLYVW 155  
Db 157 RTNSVQNRHSTFGSYFYD-----KSSSSFEIAGNEVALAIWSPNSNDIAVQDNNTIY 212  
Qy 156 ---DNGVTVRTDGGPDMFHVGDWIVYEEILGDRVALWSPDGEVLAFLSNFNETGPT 212  
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWYVEEVEFDDKAAWSPGTGDYLAFLKIDISEVGE 272  
Qy 213 YTVQYMDNQBIAPAYMELKIRYPKVSQTNPTVTLSLLNTASKEVKQAPIDAFEST-DL 271  
Db 273 FIIPYVODEK--DIYPEMRSIKYPKSGTPNPHAEALWVYSMKDGTSHPRISGNKGDSL 330  
Qy 272 IIGEVAMJTDHTTTVAAKAFNRVODQKQVAVDTASNKATVISDRDGTGDHLDNLLSMKY 331  
Db 331 LIETVTVWNGN--VLVKTTDRSSDILTFTVIDTIAKTSNVVRNESSNGGWETHNTLF 388  
Qy 332 IGPIKPSDKAY-YDISDHGSMALXLYFPVSGGEPI-PLTKGDWETS-TLSIDQERQL 388  
Db 389 IPANETFDPRPHNGYVDIILPIGGYNHLAYFENSNSHYKLTTEGKWEVNGPLAFDSMNR 448  
Qy 389 VYVLSQHHSTERHLYSVSY-STFAVTPLVDDTVAAYMSASFANSGYIILTYGGPDVPPY 447  
Db 449 LYFISTRKSTHERHYYIIDLSPNEIIIEVTDTSDDGVVDVSFSGRRRFGLLTYGPKVPY 508  
Qy 448 QELYTNTST-----KPRTTTIDNAKVLEQIKDYALPNITYPELPLPSGE----- 491  
Db 509 QKIVDFHRSKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKD 568  
Qy 492 -TLNVMQRLPGEFP--DKYPIILPTPYGGCAQEVTKRMOALNFAKVASDSSELEVVTW 548  
Db 569 ILVNSYEILPNDFETLSDHPVPVFAYGGNSQOVKTF-SVGFEVVA--SQNALIV 625  
Qy 549 TVDRNGTGKGRKFSASVTRQGLLEARDQIYAAQAAANIPIWIDADHIGIWGVSFGGYLT 608  
Db 626 VVDGRGTGFKQDFSLVDRDLRGDYEARDOIISAASLYGSLTFVDPQKISLFGWSYGYLT 685  
Qy 609 SKVLKDSGA-FTLGVIITAPVSDMRFPYSMTERTYMKTLSTNEBGETSARVKTDGFKNV 667  
Db 686 LKLTLEKDGGRHFKYGMVAPVTDWRFYDSVYTERYTMHTPQENFDGYVSSVHNVTALQA 745  
Qy 668 EGGELI OHGTCGDDNVHFONSAAVLDLLMGDGVSPKLSQWFTSDHGISYHGGGVFLYK 727  
Db 746 -NRFLMHGTGDDNVHFONSUKFLDLDLNGVENYDVHV--FPDSHDHSIRYHNANVIVFD 802  
Qy 728 QL 729  
Db 803 KL 804

RESULT 9

AD117327  
ID AD117327 standard; protein; 504 AA.

XX AC AD117327;

XX DT 15-APR-2004 (first entry)

XX DE Polypeptide homologous to a human NOVX domain SeqID 863.

XX KW NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes; inflammation;  
XX autoimmune disorder; allergy; blood disorder;  
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; infection; str.

OS Unidentified.

PT New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or

PN WO200268649-A2.  
XX 06-SEP-2002.  
XX 31-JAN-2002; 2002WO-US002785.  
XX 31-JAN-2001; 2001US-0265395P.  
PR 31-JAN-2001; 2001US-0265412P.  
PR 31-JAN-2001; 2001US-0265514P.  
PR 31-JAN-2001; 2001US-0265517P.  
PR 02-FEB-2001; 2001US-0266406P.  
PR 05-FEB-2001; 2001US-0266767P.  
PR 07-FEB-2001; 2001US-0266975P.  
PR 07-FEB-2001; 2001US-0267057P.  
PR 08-FEB-2001; 2001US-0267459P.  
PR 09-FEB-2001; 2001US-0267823P.  
PR 15-FEB-2001; 2001US-0268974P.  
PR 26-FEB-2001; 2001US-0271664P.  
PR 27-FEB-2001; 2001US-0271839P.  
PR 27-FEB-2001; 2001US-0271855P.  
PR 02-MAR-2001; 2001US-0272788P.  
PR 02-MAR-2001; 2001US-0273046P.  
PR 14-MAR-2001; 2001US-0275925P.  
PR 14-MAR-2001; 2001US-0275947P.  
PR 14-MAR-2001; 2001US-0275950P.  
PR 14-MAR-2001; 2001US-0275989P.  
PR 15-MAR-2001; 2001US-0276448P.  
PR 15-MAR-2001; 2001US-0276450P.  
PR 16-MAR-2001; 2001US-0276397P.  
PR 16-MAR-2001; 2001US-0276768P.  
PR 20-MAR-2001; 2001US-0278652P.  
PR 26-MAR-2001; 2001US-0278775P.  
PR 26-MAR-2001; 2001US-0278778P.  
PR 29-MAR-2001; 2001US-0279882P.  
PR 29-MAR-2001; 2001US-0279884P.  
PR 30-MAR-2001; 2001US-0280147P.  
PR 11-APR-2001; 2001US-0282992P.  
PR 11-APR-2001; 2001US-0283083P.  
PR 20-APR-2001; 2001US-0285133P.  
PR 23-APR-2001; 2001US-0285749P.  
PR 03-MAY-2001; 2001US-0288327P.  
PR 03-MAY-2001; 2001US-0288504P.  
PR 29-MAY-2001; 2001US-0294047P.  
PR 30-MAY-2001; 2001US-0294473P.  
PR 08-JUN-2001; 2001US-0296964P.  
PR 18-JUN-2001; 2001US-0298959P.  
PR 19-JUN-2001; 2001US-0299324P.  
PR 13-AUG-2001; 2001US-0312020P.  
PR 16-AUG-2001; 2001US-0312889P.  
PR 16-AUG-2001; 2001US-0312908P.  
PR 21-AUG-2001; 2001US-0313390P.  
PR 28-AUG-2001; 2001US-0315470P.  
PR 31-AUG-2001; 2001US-0316447P.  
PR 07-SEP-2001; 2001US-0318115P.  
PR 07-SEP-2001; 2001US-0318118P.  
PR 12-SEP-2001; 2001US-0318740P.  
PR 19-SEP-2001; 2001US-0323379P.  
PR 18-OCT-2001; 2001US-0330245P.  
PR 18-OCT-2001; 2001US-0330308P.  
PR 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.

XX Tchernev VT, Spytek KA, Zerhusen BD, Patturajan M, Shimkets RA;  
XX Li L, Gangolli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE;  
XX Gerlach VL, Taupier RJ, Gusev VV, Colman SD, Wolenc AR, Pena CEA;  
XX Furtak X, Grosse WM, Alsobrook JP, Lepley DM, Burgess CE;  
XX WPI; 2002-706998/76.

XX New NOVX polypeptides and nucleic acids, useful for preventing or  
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or

pharmacogenomics.

Disclosure; SEQ ID NO 863; 1498pp; English.

This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism. Specifically, it refers to the use of biologically active fragments for diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytostatic, cardiac, antiinflammatory, immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, antiarthritic, hepatotropic, neuroprotective, neurotropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example, neurogenesis, wound healing and angiogenesis. The nucleic acids are also used as in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. This polypeptide is a protein fragment that is homologous to a domain of a human NOVX protein of the invention.

Sequence 504 AA;

Query Match 25.3%; Score 1020; DB 5; Length 504;  
 Best Local Similarity 41.2%; Pred. No. 8.3e-75;  
 Matches 212; Conservative 100; Mismatches 180; Indels 22; Gaps 12;

QY 22 ETWVQAIPTRSRVQWLSGAEDGSVYAAEDGSLTIENIVNESRTLI PADKIPGKEA 81  
 DB 4 EDIFSFTPKYDIINWIS--DGEYLYQDDQTNLLVYVETGKT-TVLLSDTFNEFEA 59

QY 82 FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDVESLSKSVLPMPDQGB-IQYAWQSPV 140  
 DB 60 SNVLIISPDGKYILLSTNYEKRWHSYASYYIYDLNTGDVEPVAPESGNKIQYATWSPK 119

QY 141 GNTIAFVRENDLYV--WDNGTVTRITDDG-GPDMFHGVDPDWIYEEBILGDRVALWFSPDG 197  
 DB 120 GHKLAFVRDNLNLYIQKLPSGPAIQITTDGKSNDFINGIPDWIYEEBILSTDYALMWSPDG 179

QY 198 EYLAYLSFNETGVTYTVQYMDNQEIAPAYPWEKIRYPKVSQTNPTVTLNLNTA-SK 256  
 DB 180 DFLAYLRFNDSEVPVIEYPPYTDGQ-----YPDEMEIKYIPAGDPNPTVKLFVFNLDGA 235

QY 257 EVKQAPIDA-FESTDLIIGEVAMLTDTHTTAAKAFNRVQDQKVVAVDPTASNKATVISD 315  
 DB 236 SVSEIPLPASLASGDYIITFVAVTNER--LAVQWLNRDQNI SVLSLCTDTSSTWNVKN 293

QY 316 -RDGTGWLNDLSMKYIGIPKPSDKDAYIDI SDHSGWAHLVLPVSGEPTPLTKGDW 374  
 DB 294 FEDSETGWVETFPNLSL---PVFPLDGLSYLYDISDRDGYKHLAYEYLDGKEPTALTGKNW 350

QY 375 EVTSILSISOERQIVVYLSQTHSHTERHLYSVS--STFAVTVPLVDVTAAVWSASFSAN 432  
 DB 351 EVTNILGVDSKTDVFTATFEEGSGRHLIYSLSLKGKTTLSQDLSERCGIYSASFSPN 410

QY 433 SGYYILTYGGDPVPYBELYTNSTKPLRTITD NAKVLEIQIKOYALPNITTYFELPLPSGET 492  
 DB 411 AKYIILTYSGGVPVQTLHSSNDTKELRTLEDNEALKKALKNYQLPSKEGKIKLADGIT 470

QY 493 LNVWQRLPPGFGSPDKYKYPILFTYGGGGAQEVTK 526  
 DB 471 LNYQMINKPANFDFSKYPVLFVFFYGGGPGSQVTK 504

RESULT 10

AARS4612  
 ID AARS4612 standard; protein; 759 AA.  
 XX AARS4612;  
 AC AARS4612;  
 DT 25-MAR-2003 (revised)  
 DT 09-DEC-1994 (first entry)  
 DE Delta3-9 CD26.  
 XX Human; T cell activation antigen; CD26; analogues; deletion; soluble;  
 KW signal peptidase; immune-stimulating; response-stimulating; AIDS;  
 KW immunosuppression; AIDS-related complex.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT Misc-difference 2..3 /note= "Position of delta3-9 deletion"  
 XX WO9409132-A1.  
 PD 28-APR-1994.  
 PF 19-AUG-1993; 93WO-US007923.  
 PR 21-AUG-1992; 92US-00934162.  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX Morimoto C, Schlossman S, Tanaka T;  
 DR WPI; 1994-151317/18.  
 XX Polypeptide fragments and analogues of CD26 and encoding nucleic acid -  
 useful for stimulating immune response, e.g. for treatment of AIDS to  
 counteract immunosuppressive drug, and as vaccine adjuvant.  
 XX Claim 3; Page 49-52; 85pp; English.  
 XX The sequences given in AARS4612-14 represents analogues of the human T  
 cell activation antigen CD26 which have internal deletions. The analogues  
 pref. lack residues 3-9 or 24-34. These analogues are soluble under  
 physiological conditions and lack enough amino acid residues to render  
 them susceptible to cleavage by signal peptidase. The peptide fragments  
 and analogues are useful as immune or response- stimulating therapeutics,  
 eg. they may be used for treatment of disease conditions characterised by  
 immunosuppression, eg. AIDS or AIDS-related complex, other virally or  
 environmentally-induced conditions, and certain congenital immune  
 deficiencies. The peptides can be employed to increase immune function  
 which has been impaired by use of immunosuppressive drugs, such as certain  
 chemotherapeutic drugs. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 759 AA;

Query Match 24.5%; Score 987.5; DB 2; Length 759;  
 Best Local Similarity 32.6%; Pred. No. 7.6e-72;  
 Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGEGRRLTFNVTWVQKAITPTSRVQWLSGAEDGSVYAAEDGSLTIENIVNESRTLI 70  
 DB 29 TADSRKTYTLTD-YLKNTRYLKLISLRWIS--DHEVLYKQENILVFNAYEYGNSSVFLE 84

QY 71 PADKIPGKEAFNWHLPDLSVLWASNHTKQYRHSFFADYYVQDVESLSKSVLPMPDQGB- 129  
 DB 85 NSTFDFGHSINDISYSPDQGFILLEYNYVVKWRHSYASY---DIYDLNKRQLITEERI 141

QY 130 -GDIQYAWQSPVGNITAFVRENDLYV--WDNGTVTRITDDGPD-MFHGVDPDWIYEEBIL 185  
 DB 142 PNTQWTVSPVGHKLAYVWNNDIYVKEIPNLPSYRITWTGKEDIYNGITDWIYEEVF 201



```
Qy 186 GDRVALMSPDGEYLAIVLSPNETGPTVTVVQYMDNQEIAPAYPWEIKIRPKVQSQTNP 245
Db 202 SAYSALMWSNPGTFLAYAQFNDTEVPLIEYSFYSD---SLQYKTVRVYPYKAGAVNPT 258
Qy 246 VTLSLNI--ASKEVKQAPIDAFESTDLIG---EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 259 VKFFVNTDSSLVSNATSIQITAPASMLIGHYLCVDTWATQER--ISLQWLRIQNY 316
Qy 298 -QKVAVDVTASNKATVISDRD---GTDGWLNDLLSMKYIGIPKPS-----DKDAYIID 346
Db 317 VMDICDYDESSGRWNCIVARQHIEMSTTGW-----VGRFRPSEPHFTLDGNSFYKI 367
Qy 347 ISDHSQWALYLPVPSGGEPIPLTKGDWEVTSILSIDOERQIVYLSTQHHSTE--RHLY 404
Db 368 ISNEEGYRHICYFOIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 425
Qy 405 SVSSTFA-VTPL---VDDTVAAYMSASFSANSQYIITVGGPDVPVQELVTTNSTKPLR 460
Db 426 KIQLSDYTKVCLSCELNPERCQYISVSFSKEAKYIQLRCGPGGLPLYLTHSSVNDKGLR 485
Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNMQRLLPPGSPDKKYP 511
Db 486 VLENSALDKMLQNVQMPKSLDFILNETKFWYQ-----ILPPHFDKSKYPL 535
Qy 512 LFTPYGGGAQEVTKRWQALMFKAYVASDSLEYVTVTVNRGTGFGKRPKPSAVTQQLG 571
Db 536 LLDVYAGSCQKADTVFR-LNWTATLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 592
Qy 572 LLEAEDQIYAAQAAANIPWIDADHIGTWGSPGGVLTSLVKLEKDSGFTLGVTAPVSDW 631
Db 593 TFEVEDQTEARQFSKMGFVNDKRIAMGWSYGGVTSNVLGSGSGVFKCGIAVAPVSRW 652
Qy 632 RFYDSMYTERYM--KTLSTNBEGETSVA-RKTDGFKNVEGFLIQHGTGDDNVHFQNSA 688
Db 653 EYDVSVTERYMGLTPTPDNLDRYNSVTMSRAENFKQVE--YLLIHGTADDNVHFQNSA 710
Qy 689 ALVDLLMGDGVSPKHLKQWTFSDSHGISYHGGGVLYKQLARKLYQ 735
Db 711 QISKALVDVGVD---FQAMWYTDHGIASSTAHOHIYTHMSHFQIK 754

RESULT 11
AAR40909
ID AAR40909 standard; protein; 766 AA.
AC AAR40909;
XX
DE 25-MAR-2003 (revised)
DT 05-FEB-1994 (first entry)
XX
KW Sequence encoded by human CD26 cdna.
XX
OS Human T cell activation antigen; monoclonal antibody Tal.
XX
FH Homo sapiens.
XX
Key Location/Qualifiers
FT 7..28
FT /label= hydrophobic
FT 29..323
FT /label= N-terminal glycosylated region of extracellular
FT domain
FT /note= "8 sites for N-linked glycans"
FT 324..551
FT /label= Cysteine rich region of extracellular domain
FT /note= "1 N-linked glycosylation site"
FT 552..766
FT /label= C-terminal region of extracellular domain
FT /note= "1 N-linked glycosylation site & 1 catalytic site"
FT 627..631
FT /label= active site of serine protease/esterase
FT FT /note= "fits the consensus sequence GXSGX"
```

```
PN WO9316102-A1.
XX
PD 19-AUG-1993.
XX
PF 09-APR-1992; 92WO-US002892.
XX
PR 06-FEB-1992; 92US-00832211.
XX
PA (DAND ) DANA FARBER CANCER INST INC.
XX
PI Morimoto C, Schlossman SF, Tanaka T;
XX
DR WPI; 1993-272827/34.
XX
DR N-PSDB; AAQ46089.
XX
PT Polypeptide fragments of CD26 - are capable of disrupting binding of CD45
PT and CD26 and thus interfering with T-cell activation.
PS Disclosure; Page 39-43; 73pp; English.
XX
CC C26 is a human T cell activation antigen originally identified by its
CC reactivity with the MAb Tal. C26 cDNA library was constructed from human
CC PHA-activated T cells using the CDM7vector.The hydrophobic N-terminal of
CC the predicted CD26 polypeptide has the characteristics of a signal
CC sequence of the type 11 membrane protein, which is reinforced by the
CC observation that potential N-glycosylation sites are located in the
CC carboxy side of the hydrophobic core. Therefore the N-terminal 6 AAs are
CC predicted to be cytoplasmic, the next 22 AAs are predicted to transverse
CC the cytoplasmic membrane, and the 738 C-terminal AAs constitute the
CC predicted extracellular domain. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 766 AA;
Query Match 24.5%; Score 987.5; DB 2; Length 766;
Best Local Similarity 32.6%; Pred. No. 7.8e-72;
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;
Qy 11 TGEQSKLTNETVVKQAITPTSRVQWLSGAEGSYVYAAEDGSLTITENIVTHESRTLI 70
Db 36 TADSRKVTTLTD-YLKNTYRLKLYSLRWIS---DHEVLYKQENNILVFNAEYSGNSVFLE 91
Qy 71 PADKIPGKEAFNYWIHPDLSSVLWASNHTKQYRHSFPADYVQDVESLKSVPMLPDOE- 129
Db 92 NSTDFEGHSINDYSISPDGQFILLVYNYKQWRHSYTASY---DIYDLNRQLITEERI 148
Qy 130 -GDIQYQWSPVGNITAFVRENDLYV--WDNGTVTTRITDDGGPD-MFHGVDPDWIYEERIL 185
Db 149 PNTQWVTWSPVGHKLAYVWNDIYVKIEPNLPSVRITWTGKEDIYNGITDWVYEEVVF 208
Qy 186 GDRVALMSPDGEYLAIVLSPNETGPTVTVVQYMDNQEIAPAYPWEIKIRPKVQSQTNP 245
Db 209 SAYSALMWSNPGTFLAYAQFNDTEVPLIEYSFYSD---SLQYKTVRVYPYKAGAVNPT 265
Qy 246 VTLSLNI--ASKEVKQAPIDAFESTDLIG---EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 266 VKFFVNTDSSLVSNATSIQITAPASMLIGHYLCVDTWATQER--ISLQWLRIQNY 323
Qy 298 -QKVAVDVTASNKATVISDRD---GTDGWLNDLLSMKYIGIPKPS-----DKDAYIID 346
Db 324 VMDICDYDESSGRWNCIVARQHIEMSTTGW-----VGRFRPSEPHFTLDGNSFYKI 374
Qy 347 ISDHSQWALYLPVPSGGEPIPLTKGDWEVTSILSIDOERQIVYLSTQHHSTE--RHLY 404
Db 375 ISNEEGYRHICYFOIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 432
Qy 405 SVSSTFA-VTPL---VDDTVAAYMSASFSANSQYIITVGGPDVPVQELVTTNSTKPLR 460
Db 433 KIQLSDYTKVCLSCELNPERCQYISVSFSKEAKYIQLRCGPGGLPLYLTHSSVNDKGLR 492
Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNMQRLLPPGSPDKKYP 511
Db 493 VLENSALDKMLQNVQMPKSLDFILNETKFWYQ-----ILPPHFDKSKYPL 542
```



QY 512 LFTPYGGCAQVETKRWQALNFKAYVASDSELEYVTVTDNRGTGFKGRKPRSAVTROLG 571  
DB 543 LLDVYAGPCSQKADTVFR-LNWTYLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 599  
QY 572 LLEAEDQIYAAQQAANIPWIDADHIGIWMGSPGGYLTSKVLEKDSGAFTLGVTITAPVSDW 631  
DB 600 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGYVTSMLVSGSGVFKCGIAVAPVSRW 659  
QY 632 RYDSMYTRYM--KTLSTNEEGYETSAV-RKTDGFKNVEGFLIOHGTGDDNVHFQNSA 688  
DB 660 EYDSVYTRYMGLPTPEDNLDHYNSTVMSRAENFKQVE--YLLHGTADDNVHFQNSA 717  
QY 689 ALVDLLMGDVSPEKLHSQWFTSDSHGISYHGGVFLYKQLARKLYQ 735  
DB 718 QISKALVDVGD---FQAMWYTDHGHASSTAHQHIYTHMSHFQ 761

RESULT 12  
ABB08991  
ID ABB08991 standard; protein; 766 AA.  
AC ABB08991;  
XX  
XX  
DT 19-JUN-2002 (first entry)  
XX  
DE Human dipeptidyl peptidase IV.  
XX  
DE Human; dipeptidyl peptidase IV; antiasthmatic; anti-allergic;  
KW anti-inflammatory.  
XX  
XX  
OS Homo sapiens.  
XX  
PN US6337069-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 28-FEB-2001; 2001US-00794236.  
XX  
PR 28-FEB-2001; 2001US-00794236.  
XX  
PA (BMRA-) BMRA CORP BV.  
XX  
PI Grouzmann E, Lacroix J, Monod M;  
XX  
DR WPI; 2002-163235/21.  
XX  
PT Treating a patient for mucosal inflammation associated with rhinitis,  
PT sinusitis or both, by intranasally administering a peptidase that cleaves  
PT at Xaa-Pro sequences, to the patient.  
XX  
PS Disclosure; Col 9-14; 13pp; English.  
XX  
CC Thus invention relates to the treating of a patient for mucosal  
CC inflammation associated with rhinitis or sinusitis, comprising  
CC intranasally administering a peptidase. The peptidase is considered  
CC antiasthmatic, anti-allergic and anti-inflammatory in its action. The  
CC peptidase cleaves at Xaa-Pro sequences and is useful for treating a  
CC patient for mucosal inflammation associated with rhinitis or sinusitis,  
CC which is the result of allergies or asthma. This sequence represents  
CC human dipeptidyl peptidase IV  
XX  
SQ Sequence 766 AA;

Query Match 24.5%; Score 987.5; DB 5; Length 766;  
Best Local Similarity 32.6%; Pred. No. 7.8e-72;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGEGRKLTFTNETVVKQAITPSTRSVOVLSGAEDGSIYVAAEDGSLTIENIVNESRTLI 70  
DB 36 TADSRKTYTLTD-YLKNYTRKLYSLRWIS---DHEYLYKQENNILVFNAEYGNSSVFLE 91  
QY 71 PADKIPGKEAFNWIHPDLSVLWASNHTKQYRHSFFADYVYQDVESLSKSVLMPDQE- 129

DB 92 NSTFEFGHSINDYSISPDGQFILLEYNVYKQWRHSYTASY---DIYDLNKKQLTEERI 148  
QY 130 -GDIOYAQWSPVGNITAPVRENDLV--WDNGTIVTRITDDGGPD-MFHGVPDPIYEEEL 185  
DB 149 PNTQWVTSVGHKLAVWNNDIIVKLEPNLPSYRIITWTKGEDIINGITDWMVEEVF 208  
QY 186 GDYALWSPDGEYLAISFNETGVTYTVQVYMDNQEIAPAYPWLKIRYPKVSGTNP 245  
DB 209 SAYSLAWSPNGTFLAYAQFNDTEVPLIEYSFYSDE---SLQYPKTVRPYKAGAVNPT 265  
QY 246 VTLSLNT--ASKEVYKQAPIDAFESTDLIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297  
DB 266 VKFFVVTDSLSVNTNATSIQITAPASMLIGHYICDVTWATQER--ISLQWLRIQNTS 323  
QY 298 -QKVAVVDTASNKATVISDRD---GTDGWLNDLLSMKYIGPIKFS-----DKDAYYTD 346  
DB 324 VMDICDYDESSGRWNCVLAHQIEMSTTGM-----VGRFSEPHFTLDGNSFYKI 374  
QY 347 ISHSGWAHLVLPVSGGEPIPLTKGDWEVTSILSIDQERQLVYLYLSTOHHSTE--RHLY 404  
DB 375 ISNEEGYRHICYFQIDKXDKCTFTKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 432  
QY 405 SVSYSTFA-VTEPL---VDDTVAAYNSAFSANSYVILTYGGDPVYQELTYTNSTKPLR 460  
DB 433 KIQLSDYTKVTCLSCENLPERCQYYSVSFSKAKYQLRCSGPGPLTYTLHSSVNDKGLR 492  
QY 461 TITDNA--KYLEQIK-----DYALPNITYFELPLPSETLNVMLORLPPGSPDKKYPI 511  
DB 493 VLEDNSALDKMLQNVQMPKSLDFILNETKFWYQM-----ILPFPDKSKKYP 542  
QY 512 LFTPYGGCAQVETKRWQALNFKAYVASDSELEYVTVTDNRGTGFKGRKPRSAVTROLG 571  
DB 543 LLDVYAGPCSQKADTVFR-LNWTYLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 599  
QY 572 LLEAEDQIYAAQQAANIPWIDADHIGIWMGSPGGYLTSKVLEKDSGAFTLGVTITAPVSDW 631  
DB 600 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGYVTSMLVSGSGVFKCGIAVAPVSRW 659  
QY 632 RYDSMYTRYM--KTLSTNEEGYETSAV-RKTDGFKNVEGFLIOHGTGDDNVHFQNSA 688  
DB 660 EYDSVYTRYMGLPTPEDNLDHYNSTVMSRAENFKQVE--YLLHGTADDNVHFQNSA 717  
QY 689 ALVDLLMGDVSPEKLHSQWFTSDSHGISYHGGVFLYKQLARKLYQ 735  
DB 718 QISKALVDVGD---FQAMWYTDHGHASSTAHQHIYTHMSHFQ 761

RESULT 13  
AAG78417  
ID AAG78417 standard; protein; 766 AA.  
XX AAG78417;  
AC AAG78417;  
XX  
XX  
DT 12-APR-2002 (first entry)  
DE Human dipeptidyl peptidase IV amino acid sequence.  
XX  
XX 21953 prolyl oligopeptidase; antibody; proline; endopeptidase; cancer;  
KW cardiovascular disease; autoimmune disease; atopic allergy;  
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;  
KW antidiabetic; antiarthritic; antiasthmatic; anti-inflammatory;  
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;  
KW Grave's disease; neuronal disorder; demyelinating disease;  
KW dipeptidyl peptidase.  
XX  
XX Homo sapiens.  
XX  
PN WO200179473-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-US040483.







GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 02:48:49 ; Search time 90 Seconds  
(without alignments)

5918.599 Million cell updates/sec

Title: US-10-723-807-2\_COPY\_17\_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTGGSKRLTF.....EKNRQTQVLMHQWTKKLEE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3886	96.4	771	042812 ASPOR	042812 aspergillus
2	3020	74.9	773	05A242 EMENI	05A242 aspergillus
3	3010.5	74.6	765	04WRH9 ASPFU	04WRH9 aspergillus
4	3008.5	74.6	765	014425 ASPFU	014425 aspergillus
5	2610.5	64.7	775	05J6J3 TRIRU	05J6J3 trichophyto
6	2411.5	59.8	778	0512P7 MAGGR	0512P7 magnaporthe
7	2399.5	59.5	779	04HVM6 GIBZE	04HVM6 gibberella
8	1376	34.1	911	04IAY7 GIBZE	04IAY7 gibberella
9	1295	32.1	901	096VT7 ASPNG	096VT7 aspergillus
10	1289	32.0	874	07S180 EMENI	07S180 emericella
11	1284	31.8	880	05B934 EMENI	05B934 aspergillus
12	1276.5	31.7	895	07SHU8 NEUCR	07SHU8 neurospora
13	1262.5	31.3	947	04WX13 ASPFU	04WX13 aspergillus
14	1248.5	31.0	938	051XZ0 MAGGR	051XZ0 magnaporthe
15	1225.5	30.4	868	06CFX2 YARLI	06CFX2 yarrowia li
16	1193.5	29.6	883	05SH13 CRYNE	05SH13 cryptococcu
17	1193.5	29.6	883	05K721 CRYNE	05K721 cryptococcu
18	1162	28.7	793	1 YEAB SCHPO	1 YEAB schizosacch
19	1155.5	28.7	870	06BP08 DEBHA	06BP08 debaryomyce
20	1150.5	28.5	829	06CWM1 KLULA	06CWM1 kluyveromyc
21	1109	27.5	817	06FPV7 CANGA	06FPV7 candida gla
22	1098	27.2	838	05AG57 CANAL	05AG57 candida alb
23	1098	27.2	838	05AGJ4 CANAL	05AGJ4 candida alb
24	1069.5	26.5	737	0641D6 XENLA	0641D6 xenopus lae
25	1062.5	26.3	872	075C44 ASHGO	075C44 ashbya goss
26	1061	26.3	748	2 P70092 XENLA	2 P70092 xenopus lae
27	1055	26.2	818	2 Q66R87 YEAST	2 Q66R87 saccharomyc
28	1054	26.1	818	1 DAP2 YEAST	1 DAP2 saccharomyc
29	1010	25.0	759	05Z181 CHICK	05Z181 gallus gall
30	1008.5	25.0	751	2 Q75S83 AGKHB	2 Q75S83 agkistrodon
31	1007.5	25.0	751	2 Q75S82 AGKHB	2 Q75S82 agkistrodon

32 991 24.6 766 1 DPP4\_PIG  
33 987.5 24.5 766 1 DPP4\_HUMAN  
34 987.5 24.5 766 1 Q53TN1\_HUMAN  
35 984 24.4 765 1 DPP4\_BOVIN  
36 982.5 24.4 765 1 Q5R7G7\_PONPY  
37 982 24.3 760 2 Q53TP5\_HUMAN  
38 979 24.3 760 1 DPP4\_MOUSE  
39 977 24.2 760 1 SEPR\_HUMAN  
40 975 24.2 761 2 Q8R492\_RAT  
41 974.5 24.2 755 2 Q91651\_XENLA  
42 972 24.1 761 1 SEPR\_MOUSE  
43 969 24.0 765 1 DPP4\_FELCA  
44 964.5 23.9 1077 2 Q4P3P0\_USTMA  
45 962.5 23.9 878 2 Q75C14\_ASHGO

#### ALIGNMENTS

##### RESULT 1

O42812 ASPOR  
ID O42812 ASPOR PRELIMINARY; PRT; 771 AA.  
AC O42812;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Prolyl dipeptidyl peptidase precursor (EC 3.4.14.5).  
GN Name=dppIV;  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Doumas A., Van den Broek P., Affolter M., Monod M.;  
RT "Characterisation of the prolyl dipeptidyl peptidase-encoding gene (dppIV) from the koji mould Aspergillus oryzae.";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ002369; CAA05343.1; -; Genomic\_DNA.  
DR HSSP; P27487; 1PFO.  
DR MEROPS; S09.008; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000435; MIP.  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR002469; Peptidase\_S9B.  
DR InterPro; IPR000379; Ser\_estrs.  
DR Pfam; PF00930; DPPIV\_N; 1.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
DR PROSITE; PS00221; MIP; UNKNOWN\_1.  
DR HydroLase; Signal.  
FT SIGNAL 1 16 Potential.  
FT CHAIN 17 771 prolyl dipeptidyl peptidase.  
SQ SEQUENCE 771 AA; 87208 MW; 8AB0CB33797B3FC6 CRC64;

Query Match 96.4%; Score 3886; DB 2; Length 771;

Best Local Similarity 96.8%; Pred. No. 9.1e-240;

Matches 733; Conservative 1; Mismatches 19; Indels 4; Gaps 2;

QY 1 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEDSGSYVAAEDGSLTIEN 60

|||||

17 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEDGS--LVRGGRQRSHH 74

|||||

61 IVTNESRT--LIPADKIPGKEAFNFWIHPDLSSVLWASNHTKQYRHSFADYVQDVES 118

|||||

75 REHRQRVTHAHPADKIPTGKEAFNFWIHPDLSSVLWASNHTKQYRHSFADYVQDVES 134

|||||

119 LKSVPLMPDQEGDIQVAQNSPVGNTAFVRENDLYVWDNGTIVTRITDDGGPDMFHGVDPW 178

|||||

135 LKSVPLMPDQEGDIQVAQNSPVGNTAFVRENDLYVWDNGTIVTRITDDGGPDMFHGVDPW 194

QY 179 IYEEIILGDRVALWFSPPGEYLAISFNETGVTPTVVOYMDNQEIAIPAYPWEIKIRYPK 238  
DB 195 IYEEIILGDRVALWFSPPGEYLAISFNETGVTPTVVOYMDNQEIAIPAYPWEIKIRYPK 254  
QY 239 VSQTNPTVTLISLNIATSKVEKQAPIDAFESDGLIIGEVANLTDTHTTVAAKAFNRVQDOQ 298  
DB 255 VSQTNPTVTLISLNIATSKVEKQAPIDAFESDGLIIGEVANLTDTHTTVAAKAFNRVQDOQ 314  
QY 299 KVAVADTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSDKADYIIDISDHSGWAHLYL 358  
DB 315 KVAVADTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSDKADYIIDISDHSGWAHLYL 374  
QY 359 FVPSGGEPIPLTKGDWEVTSILSDOERQOLVYILSTQHSTHERHLYSVSYSTFAVTPPLVD 418  
DB 375 FVPSGGEPIPLTKGDWEVTSILSDOERQOLVYILSTQHSTHERHLYSVSYSTFAVTPPLVD 434  
QY 419 DTVAAYWASFSANSYGYILTYGGDPVPOELYTTNSTKPLRTITDNAKVLQIKDYALP 478  
DB 435 DTVAAYWASFSANSYGYILTYGGDPVPOELYTTNSTKPLRTITDNAKVLQIKDYALP 494  
QY 479 NITYFELPLPSGETLNVMLRPPGSPDKKYPILFTPYGGPGAQOEVTXKRWQALNFKAYVA 538  
DB 495 NITYFELPLPSGETLNVMLRPPGSPDKKYPILFTPYGGPGAQOEVTXKRWQALNFKAYVA 554  
QY 539 SDSELEYVTVTDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 598  
DB 555 SDSELEYVTVTDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 614  
QY 599 GWSFGGYLTSKVLKDSGAFITLVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSVA 658  
DB 615 GWSFGGYLTSKVLKDSGAFITLVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSVA 674  
QY 659 RKTDFKNVGGFLLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGISY 718  
DB 675 RKTDFKNVGGFLLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGISY 734  
QY 719 HGGGVFLYKQARKLYQEKNRQTVLMHQWTKKDLJE 755  
DB 735 HGGGVFLYKQARKLYQEKNRQTVLMHQWTKKDLJE 771

## RESULT 2

Q5A242 EMENI PRELIMINARY; PRY; 773 AA.  
AC Q5A242  
DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=AN6438.2;  
OS Aspergillus nidulans FGSC A4.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_Taxid=227321;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FGSC A4;  
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,  
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
RA Choepl Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,  
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,  
RA Gardyna S., Gerre S., Graham L., Grand-Pierre N., Hafez N.,  
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,  
RA Jaffe D., Johnson R., Jones C., Kamal M., Katat A., Karatas A.,  
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,  
RA Ma L.-J., Mabbitt P., MacLean C., Macdonald P., Major J., Manning J.,  
RA Matthews C., Mauceli E., McCarthy M., Meldrum J., Meneus L.,  
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,  
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,  
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,

RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,  
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,  
RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,  
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
RA Lander E.,  
RT "Genome Sequence of Aspergillus nidulans";  
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
CC !- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACD01000108; EAA58460.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 773 AA; 86954 MW; DEEB398491278B67 CRC64;

## Query Match 74.9%; Score 3020; DB 2; Length 773;

Beat Local Similarity 71.5%; Pred. No. 2.3e-184;

Matches 537; Conservative 107; Mismatches 101; Indels 6; Gaps 3;

QY 4 PRKPHAPTGESEKRLTFNETVVKQAIPTPSRSVOWLSGAEDGSVYAAEDGSLTIENIVT 63  
DB 20 PWRPREPRAAGSKELTFNETVISAALSPSSISVQWIAIATENDGDYVQEEDESIKIESIVT 79  
QY 64 NESRTLPADKIPGKEAFNYIHPDLSSVLWASNHTKQYRHSFFADYVYVQDVESLSKVP 123  
DB 80 NRSQTIIVPAEKIPA-DAYSIIWISPDLSAVLWATNYTKQYRHSFFADYVYVQDVETLVP 137  
QY 124 LMPQDEGDIQYAOWSPGVNTIAFYRENDLYVMDNGTVTRITDDGDPDMFHGVPMWYEE 183  
DB 138 LVEDMVGDIIQYAEWSPSGDSIAFVRGNLMTWSDGTVAITKCGGDPDMFHGVPMWYEE 197  
QY 184 ILGRYALWSPDGEYLAISFNETGVTPTVVOYMDNQEIAIPAYPWEIKIRYPKVSQTN 243  
DB 198 ILGRFALWSPDSELLAFLTFNETGVPTFTVQYFMDNQEIAIPAYPRELDIRYPKVSQTN 257  
QY 244 PTVTLISLNIATSKVEKQAPIDAFESDGLIIGEVANLTDTHTTVAAKAFNRVQDOQKVAV 303  
DB 258 PTVKLNILQUSDNTVSTIPIDVDFPSELIVGEVAVTDTHTLAVKAFNRVQDESKVIV 317  
QY 304 DTASNKATVISDRDGTGMDLNLNLSMKYIGPIKPSDKD---AYYIDISDHSGWAHLYLFP 360  
DB 318 ETASGETKIAHERDGTGMDLNLNLSISYVGLALGSGDASSAYVVDLSHSGWTHLYLFS 377  
QY 361 VSGGEPIPLTKGDWEVTSILSDOERQOLVYILSTQHSTHERHLYSVSYSTFAVTPPLVDDT 420  
DB 378 TSGGDPILPTEGEWEVTSIVSDOERELVYVLSLTHSTHERHLYSVSYRTFEITPLVDDT 437  
QY 421 VAAWASFSANSYGYILTYGGDPVPOELYTTNSTKPLRTITDNAKVLQIKDYALPNI 480  
DB 438 VEAYSVSFSKAGYIILTYAGSPVPOELYSVNVQTAPRLTLTNSAALIEKLEVALPNI 497  
QY 481 TYFELPLPSGETLNVMLRPPGSPDKKYPILFTPYGGPGAQOEVTXKRWQALNFKAYVASD 540  
DB 498 SYFELEIPSGEKLNVMLRPPGSPDKKYPILFTPYGGPGAQOEVSXKRWQSLDFNAYIASD 557  
QY 541 SELEYVTVTDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 600  
DB 558 PELEYVTVTDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 617  
QY 601 WSGGYLTSKVLKDSGAFITLVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSVAVRK 660  
DB 618 WSGGYLTSKVLKDSGAFITLVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSVAIRH 677  
QY 661 TDGFKNVVGGFLLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGISYHG 720  
DB 678 TDGFKNVVGGFLLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGIRYNG 737  
QY 721 GGVLFLYKQARKLYQEKNRQTVLMHQWTKK 751  
DB 738 GNVFLYKQARKLYQEKNRQTVLMHQWTKK 767

## RESULT 3

Q4WPH9 ASPFU  
ID Q4WPH9 ASPFU PRELIMINARY; PRT; 765 AA.  
AC Q4WPH9;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Antigenic dipeptidyl-peptidase Dpp4.  
GN ORFNames=Afu4g09320;  
OS Aspergillus fumigatus Af293.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
NCBI\_TaxID=330879;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=Af293;  
RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley,  
RA Arroyo J., Berriman M., Abe K., Archer D.B., Barrojo C., Bennett J.,  
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,  
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,  
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,  
RA Goldman G.H., Gomi K., Griffith-Jones S., Guilliam R., Haas B.,  
RA Haas H., Harris D., Horluchi H., Huang J., Humphrey S., Jimenez J.,  
RA Keller M., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,  
RA Kumagai T., Latton A., Latge J.-P., Li W., Lord A., Lu C.,  
RA Majores W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,  
RA Mouyna I., Mulligan S., Murphy L., O'Neill S., Paulsen I.,  
RA Penalva M.A., Perlea M., Price C., Pritchard B.L., Quail M.A.,  
RA Rabinowitsch E., Rawlins N., Rajandream M.-A., Reichard U.,  
RA Renaud H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,  
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,  
RA Sanchez-Perrero J.C., Saunders D., Seeger K., Squares R., Squares S.,  
RA Takeuchi M., Tekaitia F., Turner G., Vasquez de Aldana C.R., Weidman J.,  
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,  
RA Machida M., Hall N., Barrell B., Denning D.W.;  
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus  
Aspergillus fumigatus.";  
RL Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAHF01000005; EAL89855.1; -; Genomic DNA.  
SQ SEQUENCE . 765 AA; 85863 MW; DA7BF150D80D686E CRC64;

Query Match 74.6%; Score 3010.5; DB 2; Length 765;  
Best Local Similarity 72.2%; Pred. No. 9e-184;  
Matches 544; Conservative 102; Mismatches 104; Indels 3; Gaps 2;

QY 1 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVSQVLGSAEDGSYYVAEDGSLTTIEN 60  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 74  
15 IDVPQPAPYPTGSGKKRLTFNETVVKRAISPAISVEWISTSEDGDYVVQDQGLKIQS

QY 61 IVTNESRTLIPADKIPTGKEAFNFWTHPDLSVLASNHTKOYRHSFPADYVQDVESLK 120  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 132  
75 IVTNHTQTLPADKVP--EDAYSWIHPNLSSVLWATNYTKYRHSFYADYPIQDVQSMK

QY 121 SVPLMPDOEGDIOVAQWSPGVNITAFVRENDLVWDNGVTVRTTDGCPDMFGHPVDMIY 180  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 192  
133 LRPLAPDQSGDIQIAQWTPPGDAIFVRNNVFVWNASTSQITNDGCPDLFNGVPDNIY

QY 181 EEELTGRVALWFSPDGCEYLAYLSFNETGYPTTYOYVMNDQSIAPAYPWELKITRYPKVS 240  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 252  
193 EEELTGRFALWFSPDGAFLAFRLFNETGVPTFTVPMNDNEEIAPPYPRELELYPKVS

QY 241 QTNPVTLLSNTASKEVKAQPIDAPESTDLIIIGEVAMLTDDTHTVAAKFNRVQDQKV 300  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 312  
253 QTNPTVELNLELRGTERTPVPIADFADAKELIIGEAVMLTGKHVVAKFAFRNVQDQKV

QY 301 VAVDTASKATVISDRDGTGWLNDLLSMKYIGPIKPSDKDAYIIDISDSGWHLXLFPP 360  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 371  
313 VAVDVASIRSKTISRERDGTGWLNDLLSMAYIIGPIGES-KEEYIIDISQSGWAHLXLFPP

QY 361 VSGGEPLTKGDEWETSILSIDQERQLVYVLSSTQHSHTERHLYSVSYSTFAVTLPLVDDT 420  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 431  
372 VAGGEPIALTKEWEVTNWILSDIKPRQLVYFLSTKGHSTERHLYSVSWMKTEITPLVDDT

QY 421 VAAYSASFANSANGYIITYTGGDPVYPYOELVYTTSSTPLRTITDNAKVLBOIKDYALPNI 480  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 491  
432 VPVMSASFSSQCGYIILSYRGDPVYPYQDLVAINSTAPLRTITTSNAAVLNALKETYTLPLNI

QY 481 TYFELPLPSGETLNMQRLLPGCFSPDKKYPIILFTPYGGPGAQEVTKWQALLNFXYVASD 540  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 551  
492 TYFELALPSGETLNMQRLLPVKFSPPKKYPIVLFTPYGGPGAQEVSKAWQALDFKAYTASD

QY 541 SELEYVTWTDNRGTGFKGRKFRSAVTRQLGLEAEQIYAAQQAAANIPTWDADHIGTWG 600  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 611  
552 PELEYITWTDNRGTGFKGRAFCQVARSLEGELEAADQVFAAQQAALPYVDAQHAIWG

QY 601 WSFGGYLTSKULEKOSGAGFTLGIVITAPVSDWRFDVDSMYTERYMTKLSTNESEGYETSARK 660  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 671  
612 WSYGGYLTGKVIETDSGAFSLGVQTPAPVSDWRFDVDSMYTERYMTKLSTNESAAAGYNASAIRK

QY 661 TDGPKNVGEGFLIQHGTCDDNVHFNONSAAALVDLMGDCVSPKHLHSOWFTSDRDGISYHG 720  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 731  
672 VAGYKNVRGGVLIHQGTGDDNVHFNAAAALVDTLVGAGVTFPEKLQVQWFTSDHGIRYHG

QY 721 GGVLKYOLARLYQOKNRQTOVLHMHWTKKD 753  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 764  
732 GNVLRYQLSKRLYEKKRKEGEAHQSKSV

RESULT 4  
ID OT4425 ASPFU PRELIMINARY; PRT; 765 AA.  
AC OT4425;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Dipeptidyl-peptidase IV (EC 3.4.14.5).  
GN Name=dpp4;  
OS Aspergillus fumigatus (Sartorya fumigata).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
NCBI\_TaxID=5085;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CBS 144.89; TISSUE=Mycelium;  
RX MEDLINE=97378071; PubMed=9234752;  
RA Beauvais A., Monod M., Wyniger J., Debeaupuis J.P., Grouzman E.,  
RA Brach N., Svab J., Hovanessian A.G., Latge J.P.;  
RT "dipeptidyl-peptidase IV secreted by Aspergillus fumigatus, a fungus  
pathogenic to humans";  
RL Infect. Immun. 65:3042-3047(1997).  
DR EMBL; U87950; AAC34310.1; -; Genomic\_DNA.  
DR HSSP; P22411; LORW.  
DR MEROPS; S09.008; -.  
DR CO; GO:0016020; C-membrane; IEA.  
DR GO; GO:0004274; F-dipeptidyl-peptidase IV activity; IEA.  
DR GO; GO:0016787; P-hydrolase activity; IEA.  
DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.  
DR InterPro; IPRO01375; Peptidase\_S9.  
DR InterPro; IPRO02469; Peptidase\_S9B.  
DR InterPro; IPRO00379; Ser\_estrs.  
DR Pfam; PF00930; DPPIV\_N; I.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
KW Hydrolase.  
SQ SEQUENCE 765 AA; 85818 MW; 561D1CCBE71A6674 CRC64;

Query Match 74.6%; Score 3008.5; DB 2; Length 765;  
Best Local Similarity 72.4%; Pred. No. 1.2e-183;  
Matches 545; Conservative 100; Mismatches 105; Indels 3; Gaps 2;

QY 1 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVSQVLGSAEDGSYYVAEDGSLTTIEN 60  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 74  
15 IDVPQPAPYPTGSGKKRLTFNETVVKRAISPAISVEWISTSEDGDYVVQDQGLKIQS

QY 61 IVTNESRTLIPADKIPTGKEAFNFWTHPDLSVLASNHTKOYRHSFPADYVQDVESLK 120  
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 129

Db 121 SVPLMPODEGDIQVQAQSPVGNNTAFVRENDLYVMDNGTVTRIIDDGPPDMFHGVDPWIY 180  
Qy 133 LRPLAPOSGDIIQVQAQSPGTDALAFVRGNVNVFWMTNASTSQIINDGPDPLFNGVDPWIY 192  
Db 181 EEEILGRYALWFSFGDEYLAFLSFNETGVPTTYVQYMDNQEIAPAYPMELKIRYPKVS 240  
Qy 193 EEEILGRFALWFSFGDYLAFSLFNETGVPTTYVQYMDNEEIAAPPYPRLELRYPKVS 252  
Db 241 QTNPTVLSLNLINASKVQKAPIDAPESDILIEGEVWLTDTHTTVAKAFNRVODQKV 300  
Qy 253 QTNPTVELNLLELRTGERTPTPIDAFDAKELIIEGEVWLTKGHDVAVAKAFNRVODQKV 312  
Db 301 VAVDTASKATVVIDRDTGDLNLLSMKYGIPKPSDKDAYYIDISDHSGWAHLVFP 360  
Qy 313 VAVDVASLRSKTIISERDGTGDLNLLSMAYIGPIGES-KEEYIIDISDQSGWAHLVFP 371  
Db 361 VSGGEPIPLTKGDEWVTSILSIDOERQVLYYLSTQHHSTERHLYSVSYSTFAVTPLVDDT 420  
Qy 372 VAGGEPIALTKGEWVNLISIDPRQLVYFLSTKHHSTERHLYSVSMKTEITPLVDDT 431  
Db 421 VAAVMSAFSANSQYIITYGGDPVYQELVYTNSTKPLRTITDNKVLKQIKYALPNI 480  
Qy 432 VPAVMSAFSQQGGYIILSYRGDPVYQDLVAINSTAPLRTITNSAAVLNALKYTLFNI 491  
Db 481 TYFELPLSPGSLNVMQRLPGFSPDKKYPILFPTYPGGPGAQEVTKRQWALNFKAYVASD 540  
Qy 492 TYFELALPSGETLNVQMQLPVKFSPPKKYPVLFPTYPGGPGAQEVSKPQWALDFKAYIASD 551  
Db 541 SELEVVTWVNRGTGFGKGRFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWG 600  
Qy 552 PELEVITWVNRGTGYKGRFCQVASLGELEAADOVFAAQQAALPYVDAQHIAIWG 611  
Db 601 WSGGYLTSKVLKEDSGAFTLGIVITAPVSDWRFPVDSMYTERYMKTLSTNBSGYETSAVRK 660  
Qy 612 WSGGYLTGKVIETDSGAFSLGVQVAPVSDWRFPVDSMYTERYMKTLSTNBSGYETSAVRK 671  
Db 661 TDGPKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHGIGSVHG 720  
Qy 672 VAGYKXNVRGVLIIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHGIRYHG 731  
Db 721 GGVFLYKQARKLYOEKNRQTOQLMHWTKKDL 753  
Qy 732 GNVFLYRQLSKRELYEKKRKEGEAHQSKSV 764

## RESULT 5

Q5J6J3 TRIRU PRELIMINARY; PRT; 775 AA.  
ID Q5J6J3 TRIRU PRELIMINARY; PRT; 775 AA.  
AC Q5J6J3 10-MAY-2005 (TrEMBLrel. 30, Created)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Dipeptidyl-peptidase IV.  
OS Trichophyton rubrum.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;  
OC Trichophyton.  
OX NCBI\_TaxID=5551;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15632434; DOI=10.1099/mic.0.27484-0;  
RA Monod M., Lechenne B., Jousson O., Grand D., Zaugg C., Stocklin R.,  
RA Grouzmann E.;  
RT "Aminopeptidases and dipeptidyl-peptidases secreted by the  
RT dermatophyte Trichophyton rubrum."  
RL Microbiology 151:145-155 (2005).  
EMBL AY497021; AAS76665.1; -; Genomic\_DNA.  
DR GO:0016020; C:membrane; IEA.  
DR GO:0003824; F:catalytic activity; IEA.  
DR GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.

DR GO:0004185; F:serine carboxypeptidase activity; IEA.  
DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR002471; Pept\_S9\_AS.  
DR InterPro: IPR001563; Peptidase\_S10.  
DR InterPro: IPR001375; Peptidase\_S9.  
DR InterPro: IPR002469; Peptidase\_S9B.  
DR InterPro: IPR000379; Ser\_estis.  
DR Pfam: PF00930; DPIP\_N; 1.  
DR Pfam: PF00326; Peptidase\_S9; 1.  
DR PROSITE: PS00131; CARBOXYPEPT\_SER\_SER; 1.  
DR PROSITE: PS00708; PRO\_ENDOPEP\_SER; UNKNOWN\_1.  
KW Hydrolase; Transmembrane.  
SQ SEQUENCE 775 AA; 88021 MW; 157774F128C6BFDB CRC64;

Query Match 64.7%; Score 2610.5; DB 2; Length 775;

Best Local Similarity 61.6%; Pred. No. 3.5e-158;

Matches 463; Conservative 134; Mismatches 148; Indels 7; Gaps 4;

Qy 4 PRKPAPTGGSKELTFNETVVKQAITPSTRSVOVLSGAEDGSVYAAEDGSLTIENIVT 63  
Db 19 PREPRSPGTGGNKLLTYKECVPRATISPRSTSLAWINSEEDGRYISQSDGALLQNI 78  
Qy 64 NESRTLIPADKIPGKEAFNVYIHPDLSSVLWASNHTKQYRHSFPFADYVYQDVESLSVP 123  
Db 79 NTNKLVAADKVPKG--YYDYWFKPDLSAVLWATNYTKQYRHSYFANYFILDIKKSLTP 136  
Qy 124 LMPDQEGDIQVQAQSPVGNNTAFVRENDLYVMDNGTVTRIIDDGPPDMFHGVDPWIYEE 183  
Db 137 LAQDAQGDIQVQAQSPVGNNTAFVRENDLYVMDNGTVTRIIDDGPPDMFHGVDPWIYEE 196  
Qy 184 ILGRYALWFSFGDEYLAFLSFNETGVPTTYVQYMDNQEIAPAYPMELKIRYPKVSQTN 243  
Db 197 IFGRFALWFSFGDEYLAFLSFNETGVPTTYVQYMDNQEIAPAYPMELKIRYPKVSQTN 256  
Qy 244 PTVTLSLNLINASKVQKAPIDAPESDILIEGEVWLTDTHTTVAKAFNRVODQKVAV 303  
Db 257 PTVOFHLNLINASKVQKAPIDAPESDILIEGEVWLTDTHTTVAKAFNRVODQKVAV 316  
Qy 304 DTASNKATVISDRDGTGDLNLLSMKYGIPKPSDKDAYYIDISDHSGWAHLVFPVSG 363  
Db 317 KVESKESKVIREDGTGDLNLLSMKYGIPKPSDKDAYYIDISDHSGWAHLVFPVSG 373  
Qy 364 GEPTPLTKGDEWVTSILSIDOERQVLYYLSTQHHSTERHLYSVSYSTFAVTPLVDDTVA 423  
Db 374 GKEIALTKEWEVVAAILKVDTKKLIYFTSTKXSTHTRHVSYSVYDTKVMTPLVNDKEAA 433  
Qy 424 YMSAFSANSQYIITYGGDPVYQELVYTNSTKPLRTITDNKVLKQIKYALPNI 482  
Db 434 YITASFSKAGYIILSYQGNVYQELVYTNSTKPLRTITDNKVLKQIKYALPNI 493  
Qy 483 FELPLPSGETLNVQMQLPVKFSPPKKYPVLFPTYPGGPGAQEVTKRQWALNFKAYVASSE 542  
Db 494 FEIKLPSGETLNVQMQLPVKFSPPKKYPVLFPTYPGGPGAQEVTKRQWALNFKAYVASSE 553  
Qy 543 LEYVTWVNDNGTGFGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWGS 602  
Db 554 LEYVTWVNDNGTGFGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWGS 613  
Qy 603 FGGVLTSLKLEKDSGAFTLGIVITAPVSDWRFPVDSMYTERYMKTLSTNBSGYETSAVRKT 662  
Db 614 YGGFLTAKTLTSDGVTFFGISTAPVSDWRFPVDSMYTERYMKTLSTNBSGYETSAVRKT 673  
Qy 663 GFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHGIGSVHG 722  
Db 674 GFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHGIRYHG 733  
Qy 723 VFLYKQARKLYOEKNRQTOQLMHWTKKDL 753  
Db 734 TVQYKQLSKRELYEKKRKEGEAHQSKSV 765

## RESULT 6

Q51ZP7\_MAGG





RA Boukgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
RA Choepel Y., Collymore A., Cooke P., Corum B., DeArelano K.,  
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
RA Erickson J., Fero S., Ferreira P., FitzGerald M., Gage D., Galagan J.,  
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,  
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,  
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,  
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,  
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,  
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,  
RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,  
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,  
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,  
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,  
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,  
RA Talamas J., Testfaye S., Theodore J., Topham K., Travers M.,  
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
RA Lander E.;  
RT "Fusarium graminearum genome sequence."  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACM0100457; EAA75308.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 779 AA; 87757 MW; 3EC8D468C633F192 CRC64;

Query Match 59.5%; Score 2399.5; DB 2; Length 779;  
Best Local Similarity 57.0%; Pred. No. 1.1e-144;  
Matches 433; Conservative 140; Mismatches 179; Indels 7; Gaps 5;

QY 1 LDVPRKHAPTEGSGKRLTFNETVVKQAITPTSRVQVLSGAEQSGVYAAEDGSLTIE 60  
DB 20 VDPKPKHQPTGEGDRLTLTFNETSPSKALRPSSISVDWSSAGGDNVILSDGDLVLED 79  
QY 61 IVTNESTRLLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKOYRHSFPADYVQDVESLK 120  
DB 80 ILSEKTKFVTADKL P--KDLHEYVSGDAKKVLVASNTKOYRYSVPADYFVLVDSEGE 137  
QY 121 SVPLMPDQEGDIQVQWSPVGNNTAFVRENDLYV-WDNGTTRITDDGPPMFHGVDPWI 179  
DB 138 SEPLVNDQVGIQVQAFAPNGDAVAFVRDNDLFRSDGKVDQITSDGPPMFHGVDPWV 197  
QY 190 YEEETLGRYALWSPDGEYLAYLSFNETGPTVTVQYMDNOETAPAYPWEKIRYPKV 239  
DB 198 YEEETLGRYALWSPDGEYLAYLSFNETGPTVTVQYMDNOETAPAYPWEKIRYPKV 257  
QY 240 SQTNPVTLSLINTASKEVQKAPIDAFESTDLIIIEVAWLTDTHITVAAKAFNRVQDOOK 299  
DB 258 GSKNPTVELNDLVGEYKVPVPEAFPEELIIIEVAWVTKDHSALIFRAFNRVQDOQA 317  
QY 300 VVAVDTANKATVISDRDGTGWLNDLISMKGIPKSDKDAYIDISDHSWAHLVLF 359  
DB 318 HVVVNPETLKSNTVREDRDGSGLWLEHTLSISFVPLSGSGNGSYVDVNDQGMNHYLY 377  
QY 360 PVSGCEPIPLTKGDEWVTSILSIDQEROLVYLSLTOHSTERHLYSVSYSTFAPVPLVDD 419  
DB 378 PVNGGEALQLTSGEVEVSTILNDITAEKLVVFOAKRHSSTERHLYSVSYSTFAPVPLVDD 437  
QY 420 TVAAWYWSASFANSYGYIILTYGGPDVVPYQELVYTNST-KPLRTITDNAKVLQIKDYALP 478  
DB 438 SVPGYWSASFSSGGYIILHYGGPDVVPYQELVYTNST-KPLRTITDNAKVLQIKDYALP 497  
QY 479 NITYPELPLPSGETLNVORLPPGSPDKYPILETTPYGGCAQOEVTWKQALNFKAVYA 538  
DB 498 NITYPELHPDGYTLNVMQQLPPNFNASHKYPLVFTPYGGPNSQBRVAKSFQALGWKQYIS 557  
QY 539 SDESELYVTWVDNRGTGPKGKFSASVTRQLGLEAEADQIYAAQQAANI-PWIDADHIG 597  
DB 558 SDPELQYVVYVDNRGTAMRGKRYSWTSLQGLKLEPLDQIWAASEQLTSMFDPVNAEKIG 617

QY 598 IWCWSGGYLTSKVLEKDSGAFGLGVITAPVSDWRVFDVSMYTRYMYKTLSTNBEGETSA 657  
DB 618 MWGWSGGYLTAKTIADSGVFSGLITAPVSDWRVFDVSMYTRYMYKTLSTNBEGETSA 677  
QY 658 VRKTDFGNVEGGLIQHGTGDDNVHFQNSAALVLLMGDGVSPKHLHSQWFTDSDHGIS 717  
DB 678 VHDVSGFKNIAGQSVLHGTGDDNVHFQNSAALVLLMGDGVSPKHLHSQWFTDSDHGIS 737  
QY 718 YHGGGVFLYKQLARKLYQEKNRQ--TOVLMHQWTKDLE 754  
DB 738 YNGASVMIYKYLARTLYDEVKQPKAKALHTHOWNKRIE 776

RESULT 8  
Q4IAY7 GIBZE  
ID Q4IAY7 GIBZE PRELIMINARY; PRT; 911 AA.  
AC Q4IAY7;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=FG05621.1;  
OS Gibberella zeae PH-1.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
OX NCBI\_TaxID=229533;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PH-1;  
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,  
RA Boukgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
RA Choepel Y., Collymore A., Cooke P., Corum B., DeArelano K.,  
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
RA Erickson J., Fero S., Ferreira P., FitzGerald M., Gage D., Galagan J.,  
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,  
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,  
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,  
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,  
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,  
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,  
RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,  
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,  
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,  
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,  
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,  
RA Talamas J., Testfaye S., Theodore J., Topham K., Travers M.,  
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
RA Lander E.;  
RT "Fusarium graminearum genome sequence."  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACM0100031; EAA75192.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 911 AA; 102542 MW; B30DFE00EE04AB59 CRC64;

Query Match 34.1%; Score 1376; DB 2; Length 911;  
Best Local Similarity 39.9%; Pred. No. 4.1e-79;  
Matches 317; Conservative 125; Mismatches 285; Indels 68; Gaps 25;

QY 2 DVPRKHAPTEGSGKRLTFNETVVKQAITPTSRVQVLSG--AEDGSY--VYAADGSLT 57  
DB 117 DTEHDPDAATRGSGKTVTMDQ-LFDGSMQAKSHSISWAGPDGEDGLLEVGASEKPYIV 175  
QY 58 IENIVTNES-----RTLIPADKIPTGKEAFNYWIH-----PDLSVLWASNH 99  
DB 176 VEDIRKDKENRAPINTEIKASKSRKTLMD-NFFVHDGNOYTFDWNRPDPLKKVLGVNKK 234  
QY 100 TKQYRHSFFADYVYQDVESLSKSVPLMPDQ-EGDIQVQWSPVGNNTAFVRENDLYVW--- 155

Db 235 KKNWEHSAIIFILDEVTQAEPLVPGKPDARIQLATWSPKSDAVSVTQDNLLYIRRLD 294  
Qy 156 DNGVTTRITDGGPDMPHGVPMWYEEILGRYALWSPDGEYLAYSFNETGPTTV 215  
Db 295 DKQNVQITKDGPPFYFGIPDMWYEEVFGSRATWMSDEGKLSFLRTNETGPPFPV 354  
Qy 216 QYVM-----DQETAPAPVPELKIYKPKVQSTNPTVTLSLNIASKEVKQAPID-APES 268  
Db 355 QPFIKPSGTDPEKEEAYPEVQIKYKPKAGAHNPVVDLLFDYIAKQVFSVDIGAFAD 414  
Qy 269 TDLIIIEGVAWLTDTHTTAAKAFNRVQDQKQVAVDTASNKATVIS--DRDGTG-GWLDN 325  
Db 415 DRIIINLLWAGD--KAVVETNRVSDILKVVLIDVTSRKGKTNTVTDVDRIDGQWPEI 471  
Qy 326 LLSMKYI--GPKPSDKDAYIIDISHSWAHL-YLPVSGGEPPIPLTKGDWEVTSLS- 381  
Db 472 SHTYAIIPADPANGRDHOG-YVDSYLHEGYHLAYFTPLDNPDPIMLTGWEIIDDAPSA 530  
Qy 382 IDQERQVYVLTQHHSTERHLYSVSYSTFAVTPLVDDTVAAWMSASFSANSYYILTYG 441  
Db 531 IDLANNLVTFIAAKSSIQHVSVKLDGKLEALTDPEYAYDASFSKSGFVLLSYR 590  
Qy 442 GPDVPYQELYTNTST--KPLRTITDNKALEQIKOYALPNITYFELPLPSGSETLANMQL 499  
Db 591 GKVPQSVISPTSSVHDYEHVIEDNAELAEARSHELPMVKYGTLDLGDGIWVNVERR 650  
Qy 500 PPGFSPDKKYPILFTPYGPGQAEVTKRQWALNFKAYVADSELEYVTVTVNDRGTGPKG 559  
Db 651 PPHFQKKKYPVLFQOYSGFGSQSVTKRF-AVDFOQAYVA--SALGYLVVTVDPRTGFLG 707  
Qy 560 RKFPSAVTRQGLAEADQIYAAQQAANIPTWIDAHIGIWMGSGYGLTSKVLKDSG-A 618  
Db 708 RHRVVRVRLGVLGSDHIAAQAQSPASPYDAERLALWGSYGGFTLTKLEQDAGRT 767  
Qy 619 FTLGVTAPVSDWRFYDSMYTERYMKTLSTNEEGYETSARVKTGPKNVEGG---PLIOH 675  
Db 768 FSYGMAVAVPTDWRFYDSIYTERYMTPOENPGYDMSMIANATAL---GGNKRPLVMH 823  
Qy 676 GTGDDNVHVFQNSAALVDLLMGDGVSPKELHSHQWFTDSHGISYHGGGVFLYKQARKLYQ 735  
Db 824 GVADDNVHVFQNSLTLLDLSLDLAGVENYDVHV--FPDSHSHIYFHNGNRIYVDKLRNWLIN 881  
Qy 736 EKNRQTOVLHQWTK 750  
Db 882 AFN-----GEWLK 889

## RESULT 9

Q96VT7 ASPNG PRELIMINARY; PRT; 901 AA.  
AC Q96VT7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DT Dipeptidyl aminopeptidase type IV (EC 3.4.14.5).  
GN Name=dabp;  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CBS 120.49;  
RA Jalving R., ter Veen W.J., Schaap P.J.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ278532; CAC41019.1; -; Genomic\_DNA.  
DR HSSP; P27487; 1pFQ.  
DR GO; GO:0016020; C-membrane; IEA.  
DR GO; GO:0004177; F:aminopeptidase activity; IEA.  
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR002469; Peptidase\_S9B.  
DR InterPro; IPR003379; Ser\_eatrs.  
DR Pfam; PF00930; DPIP\_N; 1.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
KW Aminopeptidase; Hydrolase.  
SQ SEQUENCE 901 AA; 101256 MW; 3A519512C903D784 CRC64;  
Query Match 32.1%; Score 1295; DB 2; Length 901;  
Beat Local Similarity 38.4%; Pred. No. 6.2e-74;  
Matches 294; Conservative 133; Mismatches 287; Indels 52; Gaps 26;  
Qy 11 TGEQSKRUTFNETHVVKQAITPSTRSVQWLSG--AEDGSVYAAED---GSLTIEMI---- 61  
Db 122 SSSSGKPVTLQVLTQWL-PRSHAVSWIAGNGEDGLLVEQEGDQGGKGYLVDRSRK 180  
Qy 62 ---VTNERTLIPADKIPT-KGEAF--NYVTHPDLSSVLWASNHTKQVRHSHFADYVVD 115  
Db 181 GDATSQESRVLMEKAIVQVDGRTIPFVSTWSPNKLKULLSERENWHRHSTGKWIIFD 240  
Qy 116 VESLKSVPIMP--DQSGDIQAQWSPVGNITAFVRENDLYV--WDNGTIVTRITDDGPDMP 172  
Db 241 VATQTAQPLDPSNPGRVQLAIWSPSDMVAFVRDNNLYLRLSSKEVVPITKGGADLF 300  
Qy 173 HGVPMWYEEELGRYALWSPDGEYLAYSFNETGPTTVQYVM---DQETAPA-- 227  
Db 301 YGIPDMWYEEVFGSNVSTWMSGDGKYVAFRTNETAVPEFFVQYLYLSRPSGKRPPLGLE 360  
Qy 228 YPMELKTRYKPKVQSTNPTVTLSLNIASKEV--KOAPIDAFESTDLIIIEGVAWLTDTH 284  
Db 361 DYFVEIREIKYPKAGAPNPVMSLQFYDVEKQEVFSIEAP--DDFEDDRIVIEIW--GTSG 417  
Qy 285 TVAAKAFNRVQDQKQVAVDTASNKATV-----ISDRDGTGDLNLLSMKYIGPIKPSD 339  
Db 418 KILVPRATWRESVLKFLFDTKARTSKLVRTENVADIDG--GWERTQYTWFI-PADPSN 474  
Qy 340 KDAY--YIDISHSWAHL-YLPVSGGEPPIPLTKGDWEVTSI--LSIDERQLYVYLSIQ 395  
Db 475 GRPHDGYLDTVIHEGYEHLGYFTPLDNGSEPIILLTQGEWEVVDAPTAVDLRKGIYVFI 534  
Qy 396 HHSITERHLYSVSYSTFAVTPLVDDTVAAWMSASFSANSYVILTYGGPDVPYQELVY--T 453  
Db 535 ESPTERHLYQVNLDSGLKPLTDTSKPGYDYVFSFSGHTGYALLSVRGPSPIMQAIWNTET 594  
Qy 454 NSTKPLRTITDNKALEQIKOYALPNITYFELPLPSGSETLANMQRPLPPGSPDKKYPILF 513  
Db 595 DELKTEETIEDNAGLARWVDSYALPTEIYQNTI--DGFTLVVERRPHPHFPAKYPVLPF 653  
Qy 514 TPYGGPGAQEVTKRQWALNFKAYVADSELEYVTVTVNDRGTGFGKGRKPSRAVTRQGLLL 573  
Db 654 YLYNGPRSQTVDRKF-SIDFQSYVA--SSLGVIYVTVDGRGTGFGSKRTRCIVRGNLGY 710  
Qy 574 EAEQIYAAQQAANIPTWIDAHIGIWMGSGYGLTSKVLKDSG-AFTLGVTAPVSDWR 632  
Db 711 EAYDQITTANLWGEKPYVDERTMSWGSYGGFMTLTKLEQDAGQTFQYGMVAVPTDWR 770  
Qy 633 FYDSMYTERYMKTLSTNEEGYETSARVKTGDKFNGVEGFLIOHGTGDDNVHVFQNSAALVD 692  
Db 771 HDSIYTERYMTTPAHNPNGYNTSITDMTALQQTV--RFLVHIGASDDNVHVIQNTLVLD 829  
Qy 693 LLMGDGVSPKELHSHQWFTDSHGISYHGGGVFLYKQARKLYQAEKN 738  
Db 830 KLDLAGVQNYDLH--FYPDSHSHINFNHNAHRMVMYERLSSWLNVAFN 873

## RESULT 10

Q7SI80 EMENI PRELIMINARY; PRT; 874 AA.  
AC Q7SI80;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Dipeptidyl aminopeptidase.

```
GN Name=ste13;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RX MEDLINE=22830407; PubMed=12949156; DOI=10.1099/mic.0.0.C0119-0;
RA Dyer P.S., Paolletti M., Archer D.B.;
RT "Genomics reveals sexual secrets of Aspergillus.";
RL Microbiology 149:2301-2303(2003).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.
DR EMBL; BK001296; DAA01787.1; -; Genomic_DNA.
DR HSSP; P27487; 1NU6.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004177; F:aminopeptidase activity; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR002469; Peptidase S9B.
DR InterPro; IPR000379; Serine.
DR Pfam; PF00930; DPIP_N; I.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Aminopeptidase.
SQ SEQUENCE 874 AA; 98282 MW; 737152B6D76B6C80 CRC64;

Query Match 32.0%; Score 1289; DB 2; Length 874;
Best Local Similarity 38.2%; Pred. NO. 1.4e-73;
Matches 294; Conservative 123; Mismatches 280; Indels 72; Gaps 24;

QY 11 TGESEKRLTFNETVVKQAITPTSRVQWLSG--AEDGSYVVAED---GSLTIENI----- 61
DB 132 TSSG-KPITLQ-Q-VLSGVFLPRGHAISSVAGDGEDGLIERGEDDEAGYLRINDIROQG 189
QY 62 -----VTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFPAD 110
DB 190 KVNRLVMQKPTVGVDRITKPSATRPS-----PDLKKVLIISNQEKNRHSFTAS 239
QY 111 YVQDVESLKSVPFLMPDQ-EGDIOVAQWSVPGNTIAFVRENDLYV--WDNGTVVTRITDDG 167
DB 240 YWIFDVETQTAPLPPNNIDGRVQLWALSPKSDAIAFVRDNNLYLRKLSSEVVPIITKDG 299
QY 168 GPDMPHGVPDWTYEEIILGDRVALWFSDPGEVLAYLSFNETGVPTTYVQYMDN----- 221
DB 300 GEQLFYGVDPDWYEEVEFSGNSVTWSEGGQIAFIRTNESAVPEFPYQYFLSRPSGKKP 359
QY 222 QBIAPAYPWEIKIRYPKVSQNTPTVTLSSLNLIASKEVQAPI-DAFESTDLIIIGEVAMLT 280
DB 360 QPGLNYPEVREIKYPKAGAPNFVNLQFYDVEQGEVFSVDTDPDDFDDDRLLIEVIAWA- 418
QY 281 DTHITVAAKAFNRVQDQKVAVDVTSKATV-----ISDRDGTGDLNLLSMKYIGPI 335
DB 419 -AKGKVLVTRTNRESDLIKVLDVDESRESKLIHQIDISELDG--GWPEQTQSVRFI--- 472
QY 336 KPSDKDA-----YYIDISDHSGWAHL-YLFPVSGGEPIPLTKGDWEVTSI-LSIDQERQL 388
DB 473 -PADPDKGRPFPGYLTQTVHGGTDHLAVFTPLDNEPIMLTSGEVEVDVADPAVDLTGL 531
QY 389 VYVLSTQHHSTERHLYSYSTFAVTPVLVDDTVAAAYWSASFANSGYILTYGGDPVPYQ 448
DB 532 VYFIATKAPTERHLYRVLRLDGLTFLTDTSPQGVSVFSFGDAGYALLSYQGSIPWQ 591
QY 449 ELYTT--NSTKPLRITDPAKVLQKDYALPNITYFELPLPSGCTINVMQRLPGFSPD 506
DB 592 SIISTEGEKTTTLRIEDNTDLSKLVQAQYALPTENYQNTI-DGFTLQVVERRRPHFNPA 650
QY 507 KVPILFTPEYGGFAGQEVTKRQALNFKAYVASDSELEYVTWVDNRGTFKGRKFRSAG 566
DB 651 RKYPVLFLHYGGFGSGQTVDRRFN-VDFQSYVA--ASLGIYVTVVDRGRTGFGIGRAARCI 707
```

```
QY 567 TROLGLLEAEADQIYAAQAAANIPIWDADHIGIWGSEGGYLTSKVLEKDSG-AFTLGLVIT 625
DB 708 RGNIGHYEADIQATANKWAQKPYDSESRVAIWGSYGGFWTLKLEQDAGEFYQGMVAV 767
QY 626 APVSDMRFYDSMYTERYMKTLSTNEEGYETSVAVRKTDGFKNVGEGGFIQHGTDGDNVHFQ 685
DB 768 APVDMRFYDSVYTERYMHPTPQHNPCTGYDNTSISDMAALHN-NVRFVLVIHGASDDNVHIQ 826
QY 686 NSAAALVDLLMGDGVSPKLSQWFTSDHGHSIHGGGVFLYKQLARKLY 734
DB 827 NTLTLIDKL--DLASVQNVYDVHFYDSDHSIFFHNAHTMVYERMSPIFY 873
```

## RESULT 11

```
Q5B934 EMENI
ID Q5B934 EMENI PRELIMINARY; PRT; 880 AA.
AC Q5B934;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN2946.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArrellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Fato S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyian S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaife D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer B., Schupbach R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
RA Vasiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAC001000051; EAA63517.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 880 AA; 99015 MW; BF0FFB0CDOEA81BC CRC64;
```

```
Query Match 31.8%; Score 1284; DB 2; Length 880;
Best Local Similarity 38.4%; Pred. No. 3e-73;
Matches 293; Conservative 121; Mismatches 277; Indels 72; Gaps 24;

QY 11 TGESEKRLTFNETVVKQAITPTSRVQWLSG--AEDGSYVVAED---GSLTIENI----- 61
DB 132 TSSG-KPITLQ-Q-VLSGVFLPRGHAISSVAGDGEDGLIERGEDDEAGYLRINDIROQG 189
QY 62 -----VTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFPAD 110
DB 190 KVNRLVMQKPTVGVDRITKPSATRPS-----PDLKKVLIISNQEKNRHSFTAS 239
QY 111 YVQDVESLKSVPFLMPDQ-EGDIOVAQWSVPGNTIAFVRENDLYV--WDNGTVVTRITDDG 167
```

```
Db 240 YWFDVETQTAEPDLPNNIDGRVQLAWSPKDAIAFVRDNNLYLRLKLSERVAWPIITKDG 299
QY 168 GPMDFGVDPDWIYEELIGDRVALWFSPPGEVLAFLSNFETGVTPTVQYMDN-----221
Db 300 GQOLFVGVDPDWIYEELVSGSVTWMSDGSQAFIRTNESAVPEFFVQYFLSRSGKKP 359
QY 222 QBIAPAYPWEIKIRYPKVSQTNPVTLSLNTASKEVKQAPI-DAFESTDLIIGEVAMLT 280
Db 360 QGLENYPVRIKPKKAGAPNFVNLQYDVQGEVFSVDTPDFDDDDRLIIIEVINA- 418
QY 281 DTHTTVAAKAFNRVQDQKVAVDATSNKATV-----ISDRDGTGWLNDLNLMSKYIGPI 335
Db 419 -AGKVLVTRTNKRESILKVLVDVTESRESKLIRIQDISLSDG--CWVEPTQSVRFI--- 472
QY 336 KPSDKDA-----YYIDISHSQWHL-YLFPVSGGEPILTKGDWEVTSI-LSISOEROL 388
Db 473 -PADPKGRPPFGYLDVTVVHEGYDHLAYFTPLDNPPEPILMTSGEWEVDAPTAVDLTRGL 531
QY 389 VYVLSQHHSTHERHLYSVSYSTFAVTPLVDDTVAAYWSASFANSYGIILTYGGDPVPYQ 448
Db 532 VFIAIKEAPTEHLYRVLDSGLTPLTDTQPGYYSVSPDAGYALLSYQGSIPWQ 591
QY 449 ELYTT--NSTKPLRTITDNAKVLQIKDYALPNITYFELPSGETLNVWQRLPQGFSPD 506
Db 592 SIISTEGEKTTLRIETDNLTKSLVAQYALPTENYQNTI-DGFTLQVVERPPHFNDP 650
QY 507 KXPILPTPGGPGAGQEVTRKQWALNFKAYVASDSELEYVTVVDRNGTGFGRKRSFSAV 566
Db 651 RKPVLPHLYGGPGSQTVDRRFN-VDFQSYVA--ASLGIYVTVVDRNGTGFGRKRSFSAV 707
QY 567 TROLGLEAEADQIYAAQQAANIPIWADHIGWGSFGVLTSLKVLKDSG-AFTLGVI 625
Db 708 RGNIGHYEAIDQATAKNAQKPYDSESNALWGSYGGFWTLKLEQDAGETFGYGMVAV 767
QY 626 APVSDWRFYDSMYTERMYKTLSTNEEGYETSFAVRKTDGFKNVGEGFLIOHGTGDDNVHFO 685
Db 768 APVSDWRFYDSMYTERMYKTLSTNEEGYETSFAVRKTDGFKNVGEGFLIOHGTGDDNVHFO 826
QY 686 NSAAVLDMGDSVPEKLSHQWFTSDHIGSHVGGGVFLYKQ 728
Db 827 NTLTLIDKL--DLASQNVQDVHFPYDSDHSHSIFPHNAHTMYIER 867
```

## RESULT 12

```
Q7SHU8 NEUCR
ID Q7SHU8 NEUCR PRELIMINARY; PRT; 895 AA.
AC Q7SHU8;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU02515.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenkoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseles M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
```

```
RA Paulsen I., Sachs M.S., Lander E.S., Nusbbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABX01000002; EAA36410.1; -; Genomic_DNA.
DR HSSP; P27487; 1NU6.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR002471; Pept_S9_AS.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002469; Peptidase_S9B.
DR InterPro; IPR00379; Ser_eatra.
DR Pfam; PF00930; DPPIV_N; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; UNKNOWN 1.
KW Hydrolase; Hypothetical protein; Transmembrane.
SQ SEQUENCE 895 AA; 100692 MW; F036509E1E3D2B6D CRC64;
```

Query Match 31.7%; Score 1276.5; DB 2; Length 895;  
Best Local Similarity 38.1%; Pred. No. 9.3e-73;  
Matches 301; Conservative 134; Mismatches 281; Indels 73; Gaps 30;

```
QY 7 PHAPTEGS-KELTNETVVKQAIPTSRSVQWLSG--AEDG-----SYVAAED 53
Db 118 POATVTHSGGKVTLDQ-VLNNEWRKSHSISWAGVNGEDGLLLEKEGANKDYLVVEDV 176
QY 54 GSLLTIENVTNSTRILIPADKIPTGKEAFN--YWIHP-----DLSSVLWASNHTKQVR 104
Db 177 RAQNPSSVEASKSLI-KDKL---PEFANKTYW--PTVTVPSTRDLKKVLLATDQVNNWR 230
QY 105 HSFFADYVQDVESLKSVPIMP-DEGDIQYQWSPVGNITAFVRENDLV--WDNGTVT 161
Db 231 HSYAVYVWIFDVETQAEPLVPYDADARLQLASWSPTSDAIVVYTRDNNFLRLKLSDKIV 290
QY 162 RITDGGPDMFHGVDPDWIYEELIGDRVALWFSPPGEVLAFLSNFETGVTPTVQYMDN 219
Db 291 QITRDGSADVFGVDPDWIYEELVSGVATWMSDGNVYAFLRNTEIGTPEYFIQFVSR 350
QY 220 -DNQELIAPA---YPWELKIRYPKVSQTNPTVTLSLNLNIAKSKYKQAPIA-PESTDLIIG 274
Db 351 PSGEPEKPGEEYVPEVRQIKYKAGAHNPIVDLKFYDVKRGDVFSDISGRPADDDRLIT 410
QY 275 EVAWLTDTHTTVAAKAFNRVQDQKVAVDATSN-----KATVISDRDGTGWLNDLNLMS 329
Db 411 EVIWI---AGQVLLIKETNRVSDVMRVVLVDVGSRTGKAVRTVDVNDIDG--GWFEISHT 465
QY 330 KYIGPIKPSD--KDAYIIDISHSQWHL-YLFPVSGGEPILTKGDWEVTSI-LSISOEROL 385
Db 466 KFI-PADPANGRPDDGVDYVTHNNGDHILAYFTPLDNPIMPILMTSGDIEVVDAPSVDLQ 524
QY 386 ROLVYVLSQHHSTHERHLYSVSYSTFAVTPLVDDTVAAYWSASFANSYGIILTYGGPDV 445
Db 525 RNLVYFVTSKESIQHRHVQVKLTGDMTPVDTTSKEGYAISFSTGAGVALSVYQGPNI 584
QY 446 PYQELYTT--NSTKPLRTITDNAKVLQIKDYALPNITYFELPSGETLNVWQRLPQGF 503
Db 585 PWQKVLSTSPNPKYEHVVEENKDLAAAKKHELPINIYGTINV-DGVELNIERRPPHF 643
QY 504 SPDKKYPIILPTPGGPGAGQEVTRKQWALNFKAYVASDSELEYVTVVDRNGTGFGRKFR 563
Db 644 DKMKKYPVLFPQQYSGPVSQTVKKTFF-AVDFQSFVAAG--LGYICVTVDGRTGTFGRKFR 700
QY 564 SAVTROLGLEAEADQIYAAQQAANIPIWADHIGWGSFGVLTSLKVLKDSG-AFTLG 622
Db 701 VITRGNLGTWESHQDQIAAKHWAQKDYIDEDRLAIWGSYGGVWTLKLEQDAGETFGY 760
QY 623 VITAPVSDWRFYDSMYTERMYKTLSTNEEGYETSFAVRKTDGF-KNVEGGFLIOHGTGDDN 681
```



RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,  
RA Lindblad-Toh K., Liu X., Lokysang T., Lokysang Y., Lucien O.,  
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,  
RA Manning J., Maraballa R., Maru K., Matthews C., Mauceli E.,  
RA McCarthy M., McDonough S., McInnes T., Meldrum J., Meneus L.,  
RA Mesirov J., Mihalev A., Mihova G., Naylor J., Neves C., Nguyen C.,  
RA Mozes J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,  
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizari M., Norbu C.,  
RA Norbu N., O'donnell P., Okaewo O., O'leary S., Omotohio B.,  
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pignani B.,  
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,  
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,  
RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,  
RA Sheridan J., Shera N., Shi J., Smirnov S., Smith C., Sougnez C.,  
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,  
RA Stenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,  
RA Tensing P., Tesfaye S., Theodore J., Thouloung Y., Topham K.,  
RA Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,  
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,  
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,  
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembek L.,  
RA Zimmer A., Zody M., Lander E.;  
RT "The genome sequence of Magnaporthe grisea";  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=70-15;  
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=70-15;  
RA Zhu H., Blackmon B.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; RACU01000945; EAA53468.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE . 938 AA; 105280 MW; 394BD6652B2C8283 CRC64;

Query Match  
Best Local Similarity 31.0%; Score 1248.5; DB 2; Length 938;  
Matches 301; Conservative 116; Mismatches 304; Indels 75; Gaps 26;

QY 4 PRKHAPTGEGSKRLTFNVTVKQAITPSRSVQWLSG--AEDGSVYAAEDG-----54  
Db 135 PEKTKAAALTSKKITMDQ-VRSQWTYAAASHDITWAGPKGSDGLMIEVGADGKQYIVIE 193  
QY 55 -----SLTENVITNESRLIPADK-----IPTKEAFNWIHPDLSSVLWAS 97  
Db 194 DGRAENPDSSAAEVDVTSARLTKNGSGFRVQGRWHIFSGAEPSR-----DQCKVLLRT 248  
QY 98 NHTKQYRHFADYVQDVESLKSVPIMP-DOEGDIQYAQWSPGVNNTIAFVRENDLYVWD 156  
Db 249 EQKSNWRHSHTACYWINDVATETAEELIPASPDARVQNAQWSPSDAIVFTRENNLYRT 308  
QY 157 NGT--VTRITDGGPDMFGVDPDWIYEEBILGDRYALWSPDGEYLAYSFNETGYPTT 214  
Db 309 IGSKKVQITKDGADLFGVDPDWYEEVEIAGASTTWSDEGKYAFLETNETGVPAFP 368  
QY 215 VQYMDNQEIA-----PAYMELKIRYKVSQTNPTVTLNLINAKSKVKQAPID-APE 267  
Db 369 VQYFMSRSPGASPKAGEETPEVVDIKYPRSGSHNPVVDVQFYDIARGDVFTVDINGGA 428  
QY 268 STLIIIGEAWLTDHTTAAKAFNRVQDQKVAVD----TASNKATV-ISRDRGTDGW 322  
Db 429 DEDRLITVLV-----ACQVIVKETNRVSDIMRVLLDVKKRTGKTRTVDVGVKIDG--GW 483  
QY 323 LDNLLSMKYI--GPIKPSDKAYIIDSHSWAHL-YLPPVSGGSPILPTKG-DHEVTS 378  
Db 484 FEISQNTKYPADPKKGRDQG-YIDTVIHNDGDHLYAFSPDPNPDPFIMLTGSGWEVVD 542

QY 379 ILS-IDQRQLVYVYLSTQHHSTERHLISVSYSTFAVTLVDVDTVAAYHSASFSANSYYI 437  
Db 543 APSADVLENNLVYFIATKEGSTORHVYSVQDGNKMKSFDTTEAKGYDVVSFGAGYAL 602  
QY 438 LTYGGPDVPYQELVTT--NSKPLRTITDNAKLVLEQIKDYALPNITYFELPLPSETLNV 495  
Db 603 LSYGPKPKIPWKQVSTPANNORYEVLIEENKELASARKYDLPILNYGLTV--EGVELNY 661  
QY 496 MQRLLPPGSPDKKYPILFTPYGGGAQVETKRWQALNFKAYVASDSELEYVTTVDNRGT 555  
Db 662 LERRPPQSEKKKYPVLFHQYSGPGSQSVSKF-SVDFQSYVA--ASLGIYVTFDGRGT 718  
QY 556 GFKGKGRFSATVTRQGLLEADSDIYAAQQAANIPIWDADHIGIWCWSSGGYLTSTKLEKD 615  
Db 719 GFICRKNRVLRSRIGETIEAQDIAAKHWSLGVYDPSRIAIGWMSYGGFQTLTKLEAD 778  
QY 616 SG-AFTLGIVITAPVSDMRFYDSMYTERYMKTLSTNEEGVETSAVRKTDGFKNVEGGLIQ 674  
Db 779 AGRTFSYGMVAPVTDWRFYDSIYTERYMLTFPQONEDGYTKSAVHNSALAS-NKRFLLM 837  
QY 675 HGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQNFTSDSHGISYHGGVFLYKQARKLY 734  
Db 838 HGASDDNVHFQNSLTLKLDMGAVENYDVHV--FPDSDHSIFFHNANKIYDKLENLWV 895  
QY 735 QEKNRQTQVLMHQWTK 750  
Db 896 NAFN-----GEMLK 904

RESULT 15  
Q6CFX2 YARLI PRELIMINARY; PRT; 868 AA.  
AC Q6CFX2;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Similar to tr|Q96VT7|Aspergillus niger Dipeptidyl aminopeptidase type IV.  
GN OrderedLocusNames=YALI0B02838g;  
OS Yarrowia lipolytica (Candida lipolytica).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; Yarrowia.  
OX NCBI\_TaxID=4952;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=CLIB 122 / E 150;  
RX PubMed=15229592; DOI=10.1038/nature02579;  
RA Dujon B., Sherman D., Fischer G., Durand P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E., Barbe V.,  
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
RA Boiarane A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.-L.;  
RT "Genome evolution in yeasts.";  
RL Nature 430:35-44 (2004).  
DR EMBL; CB382128; CAG82658.1; -; Genomic\_DNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004177; F:aminopeptidase activity; IEA.  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.  
DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR002469; Peptidase\_S9B.  
DR InterPro; IPR000379; Ser\_estrs.  
DR Pfam; PF00930; DPPIV\_N; 1.







GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 03:12:40 ; Search time 27 Seconds  
(without alignments)  
2311.857 Million cell updates/sec

Title: US-10-723-807-2\_COPY\_17\_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTGGSKRLTF.....EKNRQTQLMHQWTKDLEE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5 COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/6 COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/H COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	4033	100.0	771	2	US-09-079-592-2
2	3886	96.4	771	2	US-09-462-284-2
3	1054	26.1	818	2	US-09-462-845-3
4	1054	26.1	818	2	US-10-402-312-3
5	1054	26.1	818	2	US-10-401-437-3
6	1054	26.1	818	2	US-10-402-067-3
7	1054	26.1	818	2	US-10-401-436-3
8	987.5	24.5	755	4	PCT-US93-07923-3
9	987.5	24.5	759	4	PCT-US93-07923-2
10	987.5	24.5	766	1	US-08-230-491A-3
11	987.5	24.5	766	1	US-08-619-280A-3
12	987.5	24.5	766	1	US-08-940-391-3
13	987.5	24.5	766	2	US-09-794-236-1
14	987.5	24.5	766	2	US-09-265-606-3
15	987.5	24.5	775	2	US-09-949-016-10450
16	981.5	24.3	766	2	US-10-002-593-6
17	981.5	24.3	766	2	US-09-949-016-6146
18	981.5	24.3	766	2	US-09-518-550-27
19	981.5	24.3	766	2	US-10-423-714-6
20	977	24.2	760	1	US-08-230-491A-2
21	977	24.2	760	1	US-08-619-280A-2
22	977	24.2	760	1	US-08-940-391-2
23	977	24.2	760	2	US-09-265-606-2
24	972	24.1	761	2	US-09-518-550-26
25	902.5	22.4	931	2	US-09-079-592-11
26	869.5	21.6	723	2	US-09-518-550-29
27	865.5	21.5	676	2	US-09-518-550-42

28	811.5	20.1	710	2	US-09-518-550-28	Sequence 28, Appl
29	803.5	19.9	743	2	US-10-363-937-4	Sequence 4, Appl1
30	803.5	19.9	796	2	US-09-976-674-5	Sequence 5, Appl1
31	793.5	19.7	706	2	US-09-976-674-41	Sequence 41, Appl1
32	770.5	19.1	494	2	US-09-248-796A-19472	Sequence 19472, A
33	770.5	18.7	691	2	US-09-976-674-43	Sequence 43, Appl1
34	638.5	15.8	526	2	US-09-248-796A-19924	Sequence 19924, A
35	623	15.4	732	2	US-09-518-550-30	Sequence 30, Appl1
36	613.5	15.2	593	4	PCT-US93-07923-11	Sequence 11, Appl1
37	582.5	14.4	882	2	US-09-976-674-1	Sequence 1, Appl1
38	582.5	14.4	547	2	US-10-070-464-1	Sequence 1, Appl1
39	542.5	13.5	847	2	US-09-949-016-8330	Sequence 8330, Ap
40	506.5	12.6	981	2	US-09-902-540-16812	Sequence 16812, A
41	458	11.4	863	2	US-09-976-674-3	Sequence 3, Appl1
42	458	11.4	892	2	US-09-976-674-23	Sequence 23, Appl1
43	458	11.4	892	2	US-09-976-674-27	Sequence 27, Appl1
44	456	11.3	323	2	US-09-270-767-45296	Sequence 45296, A
45	421.5	10.5	879	2	US-09-976-674-33	Sequence 33, Appl1

ALIGNMENTS

RESULT 1

US-09-079-592-2  
; Sequence 2, Application US/09079592B  
; Patent No. 6664092  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Blinkovsky  
; APPLICANT: Kimberly Brown  
; APPLICANT: Michael W. Rey  
; APPLICANT: Alan Klotz  
; APPLICANT: Tony Byun  
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl  
; FILE OF INVENTION: Amino peptidase Activity And Nucleic Acids Encoding Same  
; CURRENT APPLICATION NUMBER: US/09/079,592B  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 08/857,884  
; PRIOR FILING DATE: 1997-05-15  
; PRIOR APPLICATION NUMBER: 60/062,892  
; PRIOR FILING DATE: 1997-10-20  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 771  
; TYPE: PRT  
; ORGANISM: Aspergillus oryzae  
US-09-079-592-2

Query Match	100.0%	Score 4033;	DB 2;	Length 771;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	755;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTTSRSVQWLSGAEDSGSYVYAAEDGSLTIEN	60	
Db	17	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTTSRSVQWLSGAEDSGSYVYAAEDGSLTIEN	76	
Qy	61	IVTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYYQDVESLK	120	
Db	77	IVTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYYQDVESLK	136	
Qy	121	SVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYVWDNGTVTTRITDDGCPDMFHGVPDWIY	180	
Db	137	SVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYVWDNGTVTTRITDDGCPDMFHGVPDWIY	196	
Qy	181	EEELGDRYALWFSPDGEYLAYSFNETGVTPTTYQYMDNQEIAPAYPWLKIRYPKVS	240	
Db	197	EEELGDRYALWFSPDGEYLAYSFNETGVTPTTYQYMDNQEIAPAYPWLKIRYPKVS	256	
Qy	241	QTNPTVTLSLLNTASKEVKQAPIDAFESTDLIIIGEVANLTDTHTTVAAKAFNRVQDOQKV	300	
Db	257	QTNPTVTLSLLNTASKEVKQAPIDAFESTDLIIIGEVANLTDTHTTVAAKAFNRVQDOQKV	316	

QY	301	VAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYLFP	360
Db	317	VAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYLFP	376
QY	361	VSGGEPIPLTKGDMWEVTSILSIDOERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPLVDDT	420
Db	377	VSGGEPIPLTKGDMWEVTSILSIDOERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPLVDDT	436
QY	421	VAAVWSASFSANSNGYIILTYGGPDVPYQELVTTNSTKPLRTITDQAKVLEIQIDYALPNI	480
Db	437	VAAVWSASFSANSNGYIILTYGGPDVPYQELVTTNSTKPLRTITDQAKVLEIQIDYALPNI	496
QY	481	TYFELPLPSGETLNVMQRLPPGFPDPPKYPILFTPYGGPGAQEVTKRQALNFKAYVASD	540
Db	497	TYFELPLPSGETLNVMQRLPPGFPDPPKYPILFTPYGGPGAQEVTKRQALNFKAYVASD	556
QY	541	SELEYVTVTDNRGTGFGKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIW	600
Db	557	SELEYVTVTDNRGTGFGKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIW	616
QY	601	WSFGYLTSLVKLEKDSGFTLGVTAPVSDMRFYDSMYTERYMKTLSTNEEGYETSARK	660
Db	617	WSFGYLTSLVKLEKDSGFTLGVTAPVSDMRFYDSMYTERYMKTLSTNEEGYETSARK	676
QY	661	TGDFKNVGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTSDHGISYHG	720
Db	677	TGDFKNVGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTSDHGISYHG	736
QY	721	GGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE	755
Db	737	GGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE	771
RESULT 2			
US-09-462-284-2			
; Sequence 2, Application US/09462284			
; Patent No. 6309868			
; GENERAL INFORMATION:			
; APPLICANT: Nestec S.A.			
; APPLICANT: Monod, Michel			
; APPLICANT: Doumas, Agnes			
; APPLICANT: Affolter, Micheal			
; APPLICANT: Van Den Broek, Peter			
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM			
; TITLE OF INVENTION: CLONING OF THE			
; TITLE OF INVENTION: ASPERGILLUS ORYZAE			
; FILE REFERENCE: 8265-298			
; CURRENT APPLICATION NUMBER: US/09/462,284			
; CURRENT FILING DATE: 2000-01-03			
; NUMBER OF SEQ ID NOS: 9			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 2			
; LENGTH: 771			
; TYPE: PRT			
; ORGANISM: Fungus			
US-09-462-284-2			
Query Match 96.4%; Score 3886; DB 2; Length 771;			
Best Local Similarity 96.8%; Pred. No. 0;			
Matches 733; Conservative 1; Mismatches 19; Indels 4; Gaps 2;			
QY	1	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEADGSVYVAAEDGSLTIEN	60
Db	17	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEADGS--LRVGRQRSH	74
QY	61	IVTNESRT--LIPADKIPTGKEAFNYWIHPDLSVLWASNHTKQYRHSFFADYVVDVES	118
Db	75	REHRQVRTHAHADKIPTGKEAFNYWIHPDLSVLWASNHTKQYRHSFFADYVVDVES	134
QY	119	LKSVPLMPDQEGDIOVAQWSFVGNITAFVRENDLYVWDNGTVTRITDDGPDPMFHGVPDW	178
Db	135	LKSVPLMPDQEGDIOVAQWSFVGNITAFVRENDLYVWDNGTVTRITDDGPDPMFHGVPDW	194
f.			
QY	179	IYEEBILGDRYALWFSPDGEYLALYSFNETGVPTTYVQYTMNQEIAPAYPWELKIRYPK	238
Db	195	IYEEBILGDRYALWFSPDGEYLALYSFNETGVPTTYVQYTMNQEIAPAYPWELKIRYPK	254
QY	239	VSGTNPVTLSLNLNIAKSKVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ	298
Db	255	VSGTNPVTLSLNLNIAKSKVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ	314
QY	299	KVAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYL	358
Db	315	KVAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYL	374
QY	359	FPVSGGEPIPLTKGDMWEVTSILSIDOERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPLV	418
Db	375	FPVSGGEPIPLTKGDMWEVTSILSIDOERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPLV	434
QY	419	DTVAAYWSASFSANSNGYIILTYGGPDVPYQELVTTNSTKPLRTITDQAKVLEIQIDYALP	478
Db	435	DTVAAYWSASFSANSNGYIILTYGGPDVPYQELVTTNSTKPLRTITDQAKVLEIQIDYALP	494
QY	479	NITYFELPLPSGETLNVMQRLPPGFPDPPKYPILFTPYGGPGAQEVTKRQALNFKAYVA	538
Db	495	NITYFELPLPSGETLNVMQRLPPGFPDPPKYPILFTPYGGPGAQEVTKRQALNFKAYVA	554
QY	539	SDSELEYVTVTDNRGTGFGKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGI	598
Db	555	SDSELEYVTVTDNRGTGFGKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGI	614
QY	599	WGSFGYLTSLVKLEKDSGFTLGVTAPVSDMRFYDSMYTERYMKTLSTNEEGYETSARK	658
Db	615	WGSFGYLTSLVKLEKDSGFTLGVTAPVSDMRFYDSMYTERYMKTLSTNEEGYETSARK	674
QY	659	RKTGDFKNVGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTSDHGISY	718
Db	675	RKTGDFKNVGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTSDHGISY	734
QY	719	HGGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE	755
Db	735	HGGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE	771
RESULT 3			
US-09-462-845-3			
; Sequence 3, Application US/09462845			
; Patent No. 6723550			
; GENERAL INFORMATION:			
; APPLICANT: Estell, David A.			
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms			
; FILE REFERENCE: GC382-US			
; CURRENT APPLICATION NUMBER: US/09/462,845			
; CURRENT FILING DATE: 2000-01-13			
; PRIOR APPLICATION NUMBER: PCT/US98/14647			
; PRIOR FILING DATE: 1998-07-14			
; PRIOR APPLICATION NUMBER: EP 97305237.7			
; PRIOR FILING DATE: 1998-07-15			
; NUMBER OF SEQ ID NOS: 10			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 3			
; LENGTH: 818			
; TYPE: PRT			
; ORGANISM: Bacillus subtilis			
US-09-462-845-3			
Query Match 26.1%; Score 1054; DB 2; Length 818;			
Best Local Similarity 33.1%; Pred. No. 2e-81;			
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;			
QY	3	VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV	48
Db	48	IPHSNTPTDYOEPNSNYNDGKLVKVSFS--VVRNNTFQPKYHELOWISDNKIESNDGLLV	106
QY	49	YAAEDGSLTIENIVTNESRTLLIPADKIPTGKEAFNYWIH-----PDLSSVLW	95

Db 107 TFND-SYVVKSYVDSYNVLEGT-----FIHQNLTVESITASDLKRLI 156  
Qy 96 ASNHTKQYRHSFADYVQVDSKSLKVPMLPDQEGDIQYAOQSPVGNITAFRENDLYW 155  
Db 157 RTNSVQNRHSTFGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYQDNNIYIY 212  
Qy 156 ---DNGTVTRITDDGPDMPHGVDPDWIYEEILGDRYALWFSFDGSLAYLSFNETGVPT 212  
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWWSPTGYDLAFKIDSEVGE 272  
Qy 213 YTVQYMDQEIAPAYPWELKIRYKVSQTNPTVTLNLNIAKSEYKQAPIDAFEST-DL 271  
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTNPNAELWVYSMDKDGTSFHPRISGNKKGSL 330  
Qy 272 IIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISDRDGTGWLNDLLSMKY 331  
Db 331 LITEVTWVNGN--VLVKTDRSSDILTFLIDTIAKTSNVVRNESSNGGWEITHNTLF 388  
Qy 332 IGPIKPSDKDAY-YIDISDHSWAHLVLRPVSGEPI-PLTKGDWEVTS-ILSIDQERQL 388  
Db 389 IPANETFDPHNGYVDILPFGGYNHLAYFENSSSHYKLTTEGKWEVNGPLAFDSMENR 448  
Qy 389 VYVLSQTHSTHERHLYSVSY-STFAVTPLVDDTVAAYWSASFSANGSYVILTYGGPDVPY 447  
Db 449 LYFISTRKSTERHVVYIDLRSFNEIIEVTDSEDGVYDVSFSSGRRFGLLTYPKPKVPY 508  
Qy 448 QELYTNTST-----KPLRTITDNKVLQIKDYALPNITYFELPLPSGE----491  
Db 509 QKIVDFHSRKAECDCGNVLGSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568  
Qy 492 -TLNVQRLPPGFSP--DKKYPILFTPYGGGAEVTKRQWALNFKAYVASDSELYVTW 548  
Db 686 LKTLKDGGRHFKYGSVAPVTDWRFYDSVYTERYMHPTQENFDGYVSSVHNVTAQA 745  
Qy 688 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHGWFTDSHGSIYHGGVFLYK 727  
Db 746 -NRFLMHGTGDDNVHFQNSLKFLLDLNLGVENYDVHV--FPDSHSIRYHNANVIFD 802  
Qy 728 QL 729  
Db 803 KL 804

## RESULT 4

US-10-402-312-3  
; Sequence 3, Application US/10402312  
; Patent No. 683261  
; GENERAL INFORMATION:  
; APPLICANT: Estell, David A.  
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
; FILE REFERENCE: GC382-US  
; CURRENT APPLICATION NUMBER: US/10/402,312  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: PCT/US98/14647  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: EP 97305237.7  
; PRIOR FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-10-402-312-3

Query Match 26.1%; Score 1054; DB 2; Length 818;  
Best Local Similarity 33.1%; Pred. No. 2e-81;  
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRUTFNETVVKQAITPTSRVQMLS-----GAEDGSYV 48  
Db 48 IPHNSNTPDYQBPNSNYTNDGKLKVSFS--VVRNNTFPQPKYHELOWISDNKIESNDLGLYV 106  
Qy 49 YAAEGSLTIENIVNESRTLIPADKIPGKEAFNYIHH-----PDLSVLMV 95  
Db 107 TFND-SYVVKSYVDSYNVLEGT-----FIHQNLTVESITASDLKRLI 156  
Qy 96 ASNHTKQYRHSFADYVQVDSKSLKVPMLPDQEGDIQYAOQSPVGNITAFRENDLYW 155  
Db 157 RTNSVQNRHSTFGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYQDNNIYIY 212  
Qy 156 ---DNGTVTRITDDGPDMPHGVDPDWIYEEILGDRYALWFSFDGSLAYLSFNETGVPT 212  
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWWSPTGYDLAFKIDSEVGE 272  
Qy 213 YTVQYMDQEIAPAYPWELKIRYKVSQTNPTVTLNLNIAKSEYKQAPIDAFEST-DL 271  
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTNPNAELWVYSMDKDGTSFHPRISGNKKGSL 330  
Qy 272 IIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISDRDGTGWLNDLLSMKY 331  
Db 331 LITEVTWVNGN--VLVKTDRSSDILTFLIDTIAKTSNVVRNESSNGGWEITHNTLF 388  
Qy 332 IGPIKPSDKDAY-YIDISDHSWAHLVLRPVSGEPI-PLTKGDWEVTS-ILSIDQERQL 388  
Db 389 IPANETFDPHNGYVDILPFGGYNHLAYFENSSSHYKLTTEGKWEVNGPLAFDSMENR 448  
Qy 389 VYVLSQTHSTHERHLYSVSY-STFAVTPLVDDTVAAYWSASFSANGSYVILTYGGPDVPY 447  
Db 449 LYFISTRKSTERHVVYIDLRSFNEIIEVTDSEDGVYDVSFSSGRRFGLLTYPKPKVPY 508  
Qy 448 QELYTNTST-----KPLRTITDNKVLQIKDYALPNITYFELPLPSGE----491  
Db 509 QKIVDFHSRKAECDCGNVLGSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568  
Qy 492 -TLNVQRLPPGFSP--DKKYPILFTPYGGGAEVTKRQWALNFKAYVASDSELYVTW 548  
Db 569 ILVNSYEILPNDFDETLSDHYPVFFPAYGPNSSQVVKTF-SVGFNEVVA--SQLNAIVV 625  
Qy 549 TVDNRTGKGRKFRSAVTQGLLEAEQIYAAQQAANIPIWIDAHIGIHWGWSFGYLT 608  
Db 626 VVDGRTGFGKQDFRSLVRDLRGDYEARQIISAALYGSILTFVDPQKISLFGWSYGYLT 685  
Qy 609 SKVLEKDSGA-FTLGVITAPVSDWRFYDSMYTERYMKLTSTNEEGYETSAVRKTDGPKNV 667  
Db 686 LKTLKDGGRHFKYGSVAPVTDWRFYDSVYTERYMHPTQENFDGYVSSVHNVTAQA 745  
Qy 688 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHGWFTDSHGSIYHGGVFLYK 727  
Db 746 -NRFLMHGTGDDNVHFQNSLKFLLDLNLGVENYDVHV--FPDSHSIRYHNANVIFD 802  
Qy 728 QL 729  
Db 803 KL 804

## RESULT 5

US-10-401-437-3  
; Sequence 3, Application US/10401437  
; Patent No. 6849440  
; GENERAL INFORMATION:  
; APPLICANT: Estell, David A.  
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
; FILE REFERENCE: GC382-US  
; CURRENT APPLICATION NUMBER: US/10/401,437  
; CURRENT FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: PCT/US98/14647

```
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-401-437-3

Query Match          26.1%; Score 1054; DB 2; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-81;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

QY 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
D 48 IPHSNTPDYQEPNSNTDGLKVSFS-VVRNNTFQPKYHELOWISDNKIESNDLGLYV 106
QY 49 YAAEDGSLTIENIVTNSERTLIPADKIPTGKEAFNYWIH-----PDLSSVLW 95
D 107 TFMND-SYVVKSVYDDSYNSVLLEGT-----FIHGNQNTVESITASPDLKRLLI 156
QY 96 ASNHTKQVRHSFFADYYVQDVESLKSVPMPDQEGDIQYAOQSPVGNITAFVRENDLYW 155
D 157 RTNSQVNRHSTFGSYFYVD---KSSSFEEIGNEVALAIWSPNSNDIAYVQDNNIYI 212
QY 156 ---DNGTVTRITDDGPDPMFHGVDPMIYEELIGDRVALWFSPOGEYLAYSFNETGVPT 212
D 213 SAISKKTIRAVTNDGSSFLFNGKPDWYEEVEFDDKAAMWSPTGDYLAFLKIDSEVEGE 272
QY 213 YTVQYMDNQBIAPAYPWEKIRYPKVQSTNPTVTLSLNIASKEVKQAPIDAFEST-DL 271
D 273 FIIPYVODEK--DIYPEMRSIKYPKSGTPNPHAEALWYSMKDGTSPHPRISGNKKDGL 330
QY 272 IIGEVAMLTDRHTTTVAAKAFNRVQDQKVAVDTASNKATVISDRDGTGDLNLLSMKY 331
D 331 LITEVTWVGNGN--VLVKTDRSSDILTAVFLIDITAKTSNVVRNESSNGGWETHNTLF 388
QY 332 IGPIKPSDKAY-YTIDSDHSGWAHLXLPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388
D 389 IPANETFDPRPHNGYVDIILPIGGYNHLAYFENSNSHYKLTTEGKWEVVGPLAFDSMENR 448
QY 389 VYILSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSNGYIILTYGGPDVPY 447
D 449 LYFISTRKSTERHYYIDLRSPNEIIEVTDISEDGVYDVFSFGRRRGLLTYKGPVPY 508
QY 448 QELYTTNST-----KPLRTITDNAKVLEQIKDYALPNITYFELPLPSGE---- 491
D 509 QKIVDFHSRKAECDCGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKD 568
QY 492 -TLNVMQRLPPGFSP--DKKYPILETPPGGAQAEVTKRQWALNFKAYVASDSELEYVTW 548
D 569 ILVNSYEILPNDFDETLSDHYVPVFFAYGPGNSQQVVKTF-SVGNEVVA--SQLNAIVV 625
QY 549 TVDNRGTGKGRKFSRQVTRQGLLEAEQIYAAQQAANI PMIDADHIGIWCWSFGGYLT 608
D 626 VVDGRTGFGKQDFRSLVRDRLDGYEARDOI SAASLYGSLTFVDVQKLSLFCWSYGGYLT 685
QY 609 SKVLEKDSGA-FTLGVIITAPVSDMRFPYDSMYTERYMKTLSTNEEGYETSVAVKTKDGRNV 667
D 686 LKTLKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHPTQENFDGVSSESVHNVTALAQA 745
QY 668 EGGELIOHGTGDNVHFONSALVDLMGDGVSPEKLHSQWFTDSDHGISYHGGGVLYK 727
D 745 -NRELLMHGTGDNVHFQNSLKFLLDLLNGVENVDVHV--FPDSHSIRYHNANVIVPD 802
QY 728 QL 729
D 803 KL 804
```

```
US-10-402-067-3
; Sequence 3, Application US/10402067
; Patent No. 6881562
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/402,067
; PRIOR FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-402-067-3

Query Match          26.1%; Score 1054; DB 2; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-81;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

QY 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
D 48 IPHSNTPDYQEPNSNTDGLKVSFS-VVRNNTFQPKYHELOWISDNKIESNDLGLYV 106
QY 49 YAAEDGSLTIENIVTNSERTLIPADKIPTGKEAFNYWIH-----PDLSSVLW 95
D 107 TFMND-SYVVKSVYDDSYNSVLLEGT-----FIHGNQNTVESITASPDLKRLLI 156
QY 96 ASNHTKQVRHSFFADYYVQDVESLKSVPMPDQEGDIQYAOQSPVGNITAFVRENDLYW 155
D 157 RTNSQVNRHSTFGSYFYVD---KSSSFEEIGNEVALAIWSPNSNDIAYVQDNNIYI 212
QY 156 ---DNGTVTRITDDGPDPMFHGVDPMIYEELIGDRVALWFSPOGEYLAYSFNETGVPT 212
D 213 SAISKKTIRAVTNDGSSFLFNGKPDWYEEVEFDDKAAMWSPTGDYLAFLKIDSEVEGE 272
QY 213 YTVQYMDNQBIAPAYPWEKIRYPKVQSTNPTVTLSLNIASKEVKQAPIDAFEST-DL 271
D 273 FIIPYVODEK--DIYPEMRSIKYPKSGTPNPHAEALWYSMKDGTSPHPRISGNKKDGL 330
QY 272 IIGEVAMLTDRHTTTVAAKAFNRVQDQKVAVDTASNKATVISDRDGTGDLNLLSMKY 331
D 331 LITEVTWVGNGN--VLVKTDRSSDILTAVFLIDITAKTSNVVRNESSNGGWETHNTLF 388
QY 332 IGPIKPSDKAY-YTIDSDHSGWAHLXLPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388
D 389 IPANETFDPRPHNGYVDIILPIGGYNHLAYFENSNSHYKLTTEGKWEVVGPLAFDSMENR 448
QY 389 VYILSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSNGYIILTYGGPDVPY 447
D 449 LYFISTRKSTERHYYIDLRSPNEIIEVTDISEDGVYDVFSFGRRRGLLTYKGPVPY 508
QY 448 QELYTTNST-----KPLRTITDNAKVLEQIKDYALPNITYFELPLPSGE---- 491
D 509 QKIVDFHSRKAECDCGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDEFGKD 568
QY 492 -TLNVMQRLPPGFSP--DKKYPILETPPGGAQAEVTKRQWALNFKAYVASDSELEYVTW 548
D 569 ILVNSYEILPNDFDETLSDHYVPVFFAYGPGNSQQVVKTF-SVGNEVVA--SQLNAIVV 625
QY 549 TVDNRGTGKGRKFSRQVTRQGLLEAEQIYAAQQAANI PMIDADHIGIWCWSFGGYLT 608
D 626 VVDGRTGFGKQDFRSLVRDRLDGYEARDOI SAASLYGSLTFVDVQKLSLFCWSYGGYLT 685
QY 609 SKVLEKDSGA-FTLGVIITAPVSDMRFPYDSMYTERYMKTLSTNEEGYETSVAVKTKDGRNV 667
D 686 LKTLKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHPTQENFDGVSSESVHNVTALAQA 745
```

QY 668 EGGFLIQHGTGDDNVHFNQSAALVDLLMGDGVSPKLSHQWFTDSHGHSYHGGGVFLYK 727  
DB 746 -NRFLLMGTGDDNVHFNQSLKFLDLLDLNGVENDVHV--FPDSHRSIRYHNANVIFD 802  
QY 728 QL 729  
DB 803 KL 804  
RESULT 7  
US-10-401-436-3  
; Sequence 3, Application US/10401436  
; Patent No. 6911333  
; GENERAL INFORMATION:  
; APPLICANT: Estell, David A.  
; TITLE OF INVENTION: Processes From Gram-Positive Organisms  
; FILE REFERENCE: GC382-US  
; CURRENT APPLICATION NUMBER: US/10/401,436  
; PRIOR FILING DATE: 2003-03-26  
; PRIOR APPLICATION NUMBER: PCT/US98/14647  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: EP 97305237.7  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-10-401-436-3  
Query Match 26.1%; Score 1054; DB 2; Length 818;  
Best Local Similarity 33.1%; Pred. No. 2e-81;  
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;  
QY 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48  
DB 48 IPHNSNTDQEPNSNYNDGKLKVSFS--VRNNTFQPKYHQLQWISDNKIESNDLGLV 106  
QY 49 YAAEDGSLTIENIVNESLTPADKIPITGKEAFNWIH-----PDLSSVLW 95  
DB 107 TFMND-SYVVSVDYSVLSLEGT-----FIHQNLTVESITASPDRLKLLI 156  
QY 96 ASNHTQYRHSFADYVVOOVESLKSVPMPQDEQDIQVAKSPVGNITAFVRENDLYW 155  
DB 157 RNSVQNRHSTFGSYFVYD-----KSSSFEIENEVAIAIWPNSNDIAYQDNNIYI 212  
QY 156 ---DNGTVTRITDDGPDMPHGVDPDIYBEEIIGDRYALWFSQDGEYLAYLSFNETGVP 212  
DB 213 SAISKKTIRAVNDGSSFLFNGKPDVYEEVEPDDKAAWSPDGYLAFLKIDSEVGE 272  
QY 213 YTVQYTMNOETAPAYPMBELKIRPKVQTNPTVTLNLNIASKEVKQAPIAFEST-DL 271  
DB 273 FIPIYVQDEK--DIYPEMSIKYKSGTNPNAELWVYSMDGTSFHPRISSGNKKGSL 330  
QY 272 IIGEVAVLTDTHTTAAKAFNRVQDQKVAVDTSNKAATVLSDRDGTGDLNLLSMKY 331  
DB 331 LITEVTWVNGN--VLVKTTRSDILTVFLDITIAKTSNVVNESNGWWEITHITLF 388  
QY 332 IGPIKPSDKDAY-YIDISDSHGWAHLFLFVSGEPI-PLTKGDWEVTS-ILSIDQERQL 388  
DB 389 IPANETFDPHNGYVILPIGGYNHLAYFENSNSHYKTLTEGKEVNVNGPLAFDSWENR 448  
QY 389 VYVLSQTHSHTERHLYSVSY-STFVATPLVDDTVAAVWSAASFSANSYYILTYGGPDVY 447  
DB 449 LYFISTRKSTERHVVYIDLRSPNEIIEVTDTSDEGVDYVVSFSGRRRFGLLTYKGPVY 508  
QY 448 QELYTNTST-----KPLRITIDNAKVLQIKDVALPNITYFELPLPSG- 491  
DB 509 QKIVDFHRSKAEKCDKGNVLGKSLYHLEKNVEVLTKEILDVAVPKFSFRELNLGKDFGKD 568  
QY 492 -TLNVNQRILPPGSP--DKKYPIILFTPYGGGGAQEVTKRWQALNFKAYVASDSELYVTW 548

DB 569 ILVNSYEILPNDFBTLSDHYPVFFPFAVGPNSSQVVKTF-SVGNEVVA--SQLNAIVV 625  
QY 549 TVDNRGTGFKGRKFRSAVTRQLGLLEADQIYAAQQAANIPIWDADHIGIWMGSGGYLT 608  
DB 626 VVDGRTGFKQDFRSLVDRDLGDEYARDQISAAALYGLSLTFVDPQKISLFCWSYGGYLT 685  
QY 609 SKVLEKDSGA-FTLGVITAPVSDWRFPYDSMYTERPMKTLSTNNEBGEYETSARVKTGDFKNV 667  
DB 686 LKLTLEKGGRRHFKYGMVAPVTDWRFPYDSVYTERYMTPTQENFDGYVSSVHNVTALQA 745  
QY 668 EGGFLIQHGTGDDNVHFNQSAALVDLLMGDGVSPKLSHQWFTDSHGHSYHGGGVFLYK 727  
DB 746 -NRFLLMGTGDDNVHFNQSLKFLDLLDLNGVENDVHV--FPDSHRSIRYHNANVIFD 802  
QY 728 QL 729  
DB 803 KL 804  
RESULT 8  
PCT-US93-07923-3  
; Sequence 3, Application PC/TUS9307923  
; GENERAL INFORMATION:  
; APPLICANT: Morimoto, Chikao  
; APPLICANT: Schloesman, Stuart F.  
; APPLICANT: Tanaka, Toshiaki  
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
; SOFTWARE: WordPerfect (Version 5.0)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/07923  
; FILING DATE: 19930819  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/934,162  
; FILING DATE: 21-AUG-1992  
; APPLICATION NUMBER: 07/832,211  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00530/055002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 755  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
PCT-US93-07923-3  
Query Match 24.5%; Score 987.5; DB 4; Length 755;  
Best Local Similarity 32.6%; Pred. No. 9.1e-76;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;  
QY 11 TGEGRKLFNETVVKQAITPTSRVQWLSGAEDGSYVAAEDGSLTIENIVNESRLLI 70  
DB 25 TADSRKTYLTLD-YLQNTYRLKLYSRWIS---DHEYLYKQENILVFNAYGNSVFL 80

```
Qy 71 PADKPTGKEAFNWIHPDLSSVLWASHNTKQYRHSFPADYVQDVESLSKSVPLMPDQE- 129
Db 81 NSTDFEGHSINDYSISPDGQFILLEVNYVQWRHSYTSY---DIYDLNKRQLITEERI 137
Qy 130 -GDIQYQAOWSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MFHGVPDVIYEEIL 185
Db 138 PNTQWTVSPVGHKLAVWVNDIVVKIEPNLPSVRIITWTKEDIYNGITDWWYEEVF 197
Qy 186 GDRVALPSPDGEVLAFLSNFNETGPTVTVQYVMDNQBIAPAYPWEKIRPKVSOQNTPT 245
Db 198 SAYSALWSPNGTFLAYAQFNDTEVPLIEYSFYSD---SIQYKTVRVVPYKAGAVNPT 254
Qy 246 VTLSELLNI--ASKEVQKAPIDAFESTDLIIG-----EVAWLTDHTTVAAKAFNRVQDQ- 297
Db 255 VKFFVNTDSLSSVNTNATSIQITAPASMLIGHVLCVDTWATQER--ISLQWLRRIQNY 312
Qy 298 -QKVAVDTASNKATVISDRD-----GTDGWLNDLLSMKYIGPIKPS-----DKDAYYID 346
Db 313 VMDICDYDESSGRWNCVLAHQHIEMSITGW-----VGRFRPSEPHFTLDGNSFYKI 363
Qy 347 ISDHSGWAHLYLFPVSGGEPIPLTKGDWEVTSILSIDOERQLVYVLSQHHSTE--RHLY 404
Db 364 ISNBEGYRHICVFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 421
Qy 405 SVSVSTFA-VTPL---VDDTVAAYWSAFSANSYVILTYGGDPDVPYQELVTTNSTKPLR 460
Db 422 KIQLSDYTKVTCVLSCELNPERCQYYSVSFSKEAKYVQLRCSGPGGLPLTLHSSVNDKGLR 481
Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNVNMQRLPPGSPDKKYP 511
Db 482 VLEDNSALDKMLQNVQMPKSLDFILNETKFWYQM-----ILPPHFDKSKKPYL 531
Qy 512 LFTPYGGGAQEVTKRWQALNFKAVASDSSELYVTVTWVNRGTGFKGRKFRSAVTRQLG 571
Db 532 LLDVYAGPCSKQADTVFR-LNWTATLASTENI--IVASFDRGSGYQGDKIMHAINREL 588
Qy 572 LLEAEDQIYAAQAAANIPIWIDADHTGIWWSFGGYLTSTKVLEKSGAFTLVITAPVSDW 631
Db 589 TFEVEDQIEAARQFSGKMGFVNDKRIAIWWSYGGYVTVSMVLGSGSGVFKCAGIAPVSRW 648
Qy 632 RPYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVGEGFLIOHGTGDDNVHFQNSA 688
Db 649 EYDYSVYTERYGLPTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQOSA 706
Qy 689 ALVDLLMGDGVSPKLSHGWFTSDSHGTSYHGGGVFLYQLARKLYQ 735
Db 707 QISKALVDVGD---FQAWWYTDHGHASSTAHQHIYTHSHFIKQ 750
```

```
RESULT 9
PCT-US93-07923-2
; Sequence 2: Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schloesman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
```

```
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US93-07923-2
;
; Query Match 24.5%; Score 987.5; DB 4; Length 759;
; Best Local Similarity 32.6%; Pred. No. 9.2e-76;
; Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;
;
; Qy 11 TGEBSKRLTFNETHVVKQAITPSTRSVQWLSGABSGSYVYAAEDGSLTIENIVTHESRTLI 70
; Db 29 TADSRKVTYLTLD-YLKNTRYRLKLYSLRWIS---DHEYLKQENNILVFNAEYGNSSVFLE 84
; Qy 71 PADKPTGKEAFNWIHPDLSSVLWASHNTKQYRHSFPADYVQDVESLSKSVPLMPDQE- 129
; Db 85 NSTDFEGHSINDYSISPDGQFILLEVNYVQWRHSYTSY---DIYDLNKRQLITEERI 141
; Qy 130 -GDIQYQAOWSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MFHGVPDVIYEEIL 185
; Db 142 PNTQWTVSPVGHKLAVWVNDIVVKIEPNLPSVRIITWTKEDIYNGITDWWYEEVF 201
; Qy 186 GDRVALPSPDGEVLAFLSNFNETGPTVTVQYVMDNQBIAPAYPWEKIRPKVSOQNTPT 245
; Db 202 SAYSALWSPNGTFLAYAQFNDTEVPLIEYSFYSD---SIQYKTVRVVPYKAGAVNPT 258
; Qy 246 VTLSELLNI--ASKEVQKAPIDAFESTDLIIG-----EVAWLTDHTTVAAKAFNRVQDQ- 297
; Db 255 VKFFVNTDSLSSVNTNATSIQITAPASMLIGHVLCVDTWATQER--ISLQWLRRIQNY 316
; Qy 298 -QKVAVDTASNKATVISDRD-----GTDGWLNDLLSMKYIGPIKPS-----DKDAYYID 346
; Db 313 VMDICDYDESSGRWNCVLAHQHIEMSITGW-----VGRFRPSEPHFTLDGNSFYKI 367
; Qy 347 ISDHSGWAHLYLFPVSGGEPIPLTKGDWEVTSILSIDOERQLVYVLSQHHSTE--RHLY 404
; Db 368 ISNBEGYRHICVFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 425
; Qy 405 SVSVSTFA-VTPL---VDDTVAAYWSAFSANSYVILTYGGDPDVPYQELVTTNSTKPLR 460
; Db 426 KIQLSDYTKVTCVLSCELNPERCQYYSVSFSKEAKYVQLRCSGPGGLPLTLHSSVNDKGLR 485
; Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNVNMQRLPPGSPDKKYP 511
; Db 486 VLEDNSALDKMLQNVQMPKSLDFILNETKFWYQM-----ILPPHFDKSKKPYL 535
; Qy 512 LFTPYGGGAQEVTKRWQALNFKAVASDSSELYVTVTWVNRGTGFKGRKFRSAVTRQLG 571
; Db 536 LLDVYAGPCSKQADTVFR-LNWTATLASTENI--IVASFDRGSGYQGDKIMHAINREL 592
; Qy 572 LLEAEDQIYAAQAAANIPIWIDADHTGIWWSFGGYLTSTKVLEKSGAFTLVITAPVSDW 631
; Db 589 TFEVEDQIEAARQFSGKMGFVNDKRIAIWWSYGGYVTVSMVLGSGSGVFKCAGIAPVSRW 652
; Qy 632 RPYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVGEGFLIOHGTGDDNVHFQNSA 688
; Db 653 EYDYSVYTERYGLPTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQOSA 710
```

```
QY 689 ALVDLLMGDVSPEKLHDSQWFTSDHGISYHGSGVFLYKQARKLYQ 735
Db 711 QISKALVDVGVD---FQAWYTDDEHGIASSTAHQHIYTHMSHFQ 754

RESULT 10
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FEFEL & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 838-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-230-491A-3

Query Match 24.5%; Score 987.5; DB 1; Length 766;
Best Local Similarity 32.6%; Pred. No. 9.3e-76;
Matches 250; Conservative 111; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGGGKRLTFNFTVVKQAITPTSRVQWLSGADGSGYVVAEDGSLTIENIVTNSRTLI 70
Db 36 TADSRKTYLTLD-YLKNYRLKLYSLRWIS--DHEYLYKQENNILVFAEYGNSSVFLE 91
QY 71 PADKIFGTGKEAFNWIHPDLSSVLWASNHTKQYRHSFPADYVQDVESLKSVPMPDQE- 129
Db 92 NSTFDFEGHSINDYSPDGQFILLVYVVKQWRHSYTSY---DIYDLNKRQLITEERI 148
QY 130 -GDIQVQWSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MFHGVDPDWIYEEIL 185
Db 149 PNTQVTVWSPVGHKLAYVWNNDIYKIEBFLPSYRITWTGKEDIYNGITDWWYEEVF 208
QY 186 GDRYALWSPDGBYLAYSFNFTGVTYVQYVYMDNQEIAPYVWELKIRYKVSQTNPT 245
Db 209 SAYSAWLSWSPNGTFLAYAQENDTEVPLIEYSFYSDE---SLQPKTVRVPPYKAGAVNPT 265
QY 246 VTLSLNI--ASKEVQAPIDAFESTDLIIG-----EVWLTDTHTTTVAAKAFNRVQDQ- 297
Db 266 VKPFVNTDLSVNTWATNSQITAPASMLIGDHYLCVDVWATQER--ISLQWLRRIQNS 323
QY 298 -QKVAVADVATSNKATVTSRDRD---GTDGWLNDLLSMKYGIPKPS-----DKDAYIID 346
; US-08-230-491A-3

US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-619-280A-3

Db 324 VMDICDYDESSGRWNCLVARQHIEMSTTGW-----VGRFRSEPEHFTLDGNSFYKI 374
QY 347 ISDHSGWAHLVLPVSGGEPILTKGDMEVTSILSDQERQLVYLYLSTQHHSTE--RHLY 404
Db 375 ISNEEGYRHCYFQIDKKDCTFITGTWEVIGIEALTS--LYYISNEYKMGPGGRNLY 432
QY 405 SVSYSTFA-VTPL-----VDDTVAAYWSAFSANSYVILTYGGDPVYQYLYTNTSKPLR 460
Db 433 KIQLSDYTKVTCLSCENLPERCQYYSVSFSEAKTYQLRCSGGLPLTYLHSSVNDKGLR 492
QY 461 TITDNA---KYLEQIK-----DYALPNITYFELPLPSGETLNVQRLPSPGSPDKKYP 511
Db 493 VLEDNSALDKMLQNMPSKSLDFILNETKFWQM-----ILPHFDKSKKPYL 542
QY 512 LFTPYGGGCAQEVTKRQALNFKAYVASDSELEYVTWTVDNRGTGFKGRKFRSAVTRQLG 571
Db 543 LLDVYAGPCSKADTVFR-LNWAYTLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 599
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGIMGWSFGGYLTSKVKLEKDSGAFITGVITAPVSDW 631
Db 600 TPEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGYVTSMLVSGSGVFKCGIAVAPVSRW 659
QY 632 RPYDSMYTERYM--KTLSTNEGYETSAV-RKTDGFKNVGEGFLIOHGTGDDNVHFQNSA 688
Db 660 EYDYSYTERYMGFLTPEDNLDHYNSTVNSRAENFKQVE--YLLHGTADDNVHFQNSA 717
QY 689 ALVDLLMGDVSPEKLHDSQWFTSDHGISYHGSGVFLYKQARKLYQ 735
Db 718 QISKALVDVGVD---FQAWYTDDEHGIASSTAHQHIYTHMSHFQ 761
```

Query Match	24.5%	Score 987.5	DB 1	Length 766
Best Local Similarity	32.6%	Pred. No. 9.3e-76		
Matches 250	Conservative 131	Mismatches 303	Indels 83	Gaps 27
11	TGEGSKRLTFNETVVKQAITPTSRVOWLSGAEDGSVYAAEDGSLTIENIVTNESRTLI	70		
36	TADSRKTYTLTD-VLKNTYRLKLYSLRWIS---DHELYKQENNILVFNAYEYGNSSVFLE	91		
71	PADKIPTGKEAFNWIHPDLSSVLWASNHTKQYRHSFADYVYQDVESLKSVPMPDQD-	129		
92	NSTDFEGHSINDYSISPDGQFILLEYNVYKQMRHSYTSY---DIYDLNKRQLITEERI	148		
130	-GDIQYAOQSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MPHGVPDWTYEEEL	185		
149	PNNQWTVWSPVGHKLAVWNNDIYVKEPLPSYRITWTGKEDIINGITDWTYEESEVF	208		
186	GDRYALWSPDGEVYLAYSFNETGVPTVYQYVMDNQEIAFPAYPWEIKIRYPKVSQTNPT	245		
209	SAYSALWSPNGTFLAYAQFNDTEVPLIEYSFYSDE---SLQYKTVRVPYKAGAVNPT	265		
246	VTLSLNI--ASKEVKQAPIDAFSTDLIIG-----EVAMLTDTHTTVAAKAFNRVQDQ-	297		
266	VKFFVNTDSSLSSVTNATSIQITAPASMLIGDHVLCDDVTWATQER--ISLQWLRRIQNY	323		
298	-QKVAVDTASNKATVISDRD-----GTDGWLNDLLSMKYIGPKPS-----DKDAYVID	346		
324	VMDICDYDESSGRWCLVARQHIENSTTGW-----VGRFPSEPHFTLDGNSFYKI	374		
347	ISDHSGWAHLVFPVSGGEPILTKGDWEVTSILSIDOERQLVYVLTQHHSTE--RHLY	404		
375	ISNEEGYRHCYFQDKKDCFTFKGTWEVIGIEALTSQD--YLYISNEYKMGPGGRNLY	432		
405	SVSYSTFA-VTPEL---VDDTVAAYWASFSANSYGIIITYGGDPVPOELYTTSKPLR	460		
433	KIQLSDYTKVTKLSCELNPERCQYYSVSFSKEAKYQLRCSGPGPLPLYTLHSSVNDKGLR	492		
461	TTTDNA---KVLQK-----DYALPNITVFELPLPSGETLNWQRLPPGSPDKKPYI	511		
493	VLEDNSALDKMLQNVQMPFSKKLDFILNETFPWQM-----ILPPHFDKSKKPYL	542		
512	LFTPYGGCAQEVTKWQALNPKAVVSDSELYVTVTDNRGTGFKGRKFRSAVTRQLG	571		
543	LLDVYAGPCSQKADTVFR-LNWTATLASTENI--IVASFDRGSGYQGDKIMHAINRRLG	599		
572	LLEAEDQYAAQAAANIPIWADHIGIWGSGFGLTSLKLEKDSGAFTLGVITAPVSDW	631		
600	TFEVDQIEAARQFQKMGFVNDKRIAIWGSYGGVTSWLGSGSGVFKGIAVAPVSRW	659		
632	RFVDSMYTERY--KTLSTNEGYETSAV-RKTDGFKNVEGFLIOHGTGDDNVHFQNSA	688		
660	EYDSVYTERYMGFLTPEDNLDHYNSIVMSRAENFKQVE--YLLIHGTADDNVHFQNSA	717		
689	ALVDLLMGDGVSEPKLHSGQWFTSDSHGISYHGGGVFLYQLARKLYQ	735		
718	QISKALVDVGDV---FQAMWYTDDEHGIIASSTAHQHIYTHMSHFQIKQ	761		

RESULT 12

US-08-940-391-3

Sequence 3, Application US/08940391

Patent No. 5965373

GENERAL INFORMATION:

APPLICANT: Zimmermann, Rainer; Park, John E.;

APPLICANT: Rettig, Wolfgang; Old, Lloyd J.

TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION

TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Felle & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York



QY 572 LLEAEDQIYAAQAAANIPWIDADHIGWGSFGYLTSLVKLEKSGAFTLGVITAPVSDW 631  
DB 600 TFEVEDQTEAARQFSKMGFVNDKRIAIWGSYGGYVTSWVLGSGGVFKCGIAPVSRW 659  
QY 632 RPYDSMYTERYM--KTLSTNEEGYETS AV-RKTDGFKNVEGFLIQHGTGDDNVHFQNSA 688  
DB 660 EYDSVYTERYMGFLTPEDNLDHYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717  
QY 689 ALVDLLMGDGVSPKHLHSQWFTSDHGISYHGGGVFLYKQLARKUYQ 735  
DB 718 QISKALVDVGD---FQAMWYTDHGIASSTAHQHIYTHMSHFQK 761

RESULT 13  
US-09-794-236-1  
; Sequence 1, Application US/09794236  
; Patent No. 6337069  
; GENERAL INFORMATION:  
; APPLICANT: Grouzmann, Eric  
; APPLICANT: Lacroix, Jean-Silvain  
; APPLICANT: Monod, Michel  
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis  
; FILE REFERENCE: 81985/276823  
; CURRENT APPLICATION NUMBER: US/09/794,236  
; CURRENT FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 766  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-794-236-1

Query Match 24.5%; Score 987.5; DB 2; Length 766;  
Best Local Similarity 32.6%; Pred. No. 9.3e-76;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TEGSGKRLTFNVTWKQAITPTSRVQWLSGAEDESYVYAAEDGSLTIENIVTNSRTLI 70  
DB 36 TADSRKTYLTLD-YLKNYRLKLYSLRWIS---DHEYLKQENNLVFNFAEYGNSSVFLE 91  
QY 71 PADKIPTGKEAFNYWIHPDLSSVWASNHTKQYRHSFFADYVQVDESLSKSVPLPDOOB- 129  
DB 92 NSTDEFHGSINDYISPDGQFILLIENYVWKWRHSYTSY---DIYDLNKKQLITEERI 148  
QY 130 -GDIQVQWSPGNTIAFVRENDLYV--WDNGTVTRITDDGGPD-MFHGVPDVIYEEEL 185  
DB 149 PNTQWTVWSPVGHKLAVVWNDIYVKIEPNLPSYRITWTGKEDIYNGITDWMVYEEVF 208  
QY 186 GDRYALWSPDGEYLAYLSFNETGVPTTYVQYMDNQEATAPWELKIRYKVSQTNPT 245  
DB 209 SAYSAWSPNGTFLAYAQFNDTEVPLEIYFSYSD---SLQYPKTVRPVYPKAGAVNPT 265  
QY 246 VTLSLNI--ASKEVKAQPIDAFESTDLIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297  
DB 266 VKFFVNTDLSVTVNATSIQTAPASMLIGHYLCVDTWATQER---ISLOWLRILQYVS 323  
QY 298 -QKWAVDTSNATKATVISDRD---GTDGWLNDLSMKYIGPIKPS-----DKDAYIID 346  
DB 324 VMDICDYDESSGRWNCVLAHQHIMSTTGM-----VGRFRPSEPHTLDGNSFYKI 374  
QY 347 ISDHGWAHLXYLPVWGGSPILTKGDWEVTSILSDQERQLVYVLSQHHSTE--RHLY 404  
DB 375 ISNEGYRHCYFQIDKCDCTTKGTWIEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 432  
QY 405 SVSYSTFA-VTFL---VDDTVAAVWASPSANSYVILTYGGDPVQYQELVYTNSTKPLR 460  
DB 433 KIQLSDYTKVTCLCSLNPFCOQYVSFSKEAKYVQLRCSGGLPLTYLTHSSVNDKGLR 492  
QY 461 TITDNA---KVLBOIK-----DYALPNITYFELPLPSGETLNVMQRLPPGFSPPDKYPI 511  
DB 493 VLEDNSALDKMLQVQWMPSKULDPIILNETKFWQM-----ILPPHFDKSKYPL 542

QY 512 LFTPYGPGPAQEVTKRQWALNFKAYVASDSELEYVVTWVDRGTGFKGKFRSAVTRQIG 571  
DB 543 LLDVYAGPCSKQADTVFR-LNWAYTLASTENI--IVASFDCGRGSGYQGDKIMHAINRRLG 599  
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGWGSFGYLTSLVKLEKSGAFTLGVITAPVSDW 631  
DB 600 TFEVEDQTEAARQFSKMGFVNDKRIAIWGSYGGYVTSWVLGSGGVFKCGIAPVSRW 659  
QY 632 RPYDSMYTERYM--KTLSTNEEGYETS AV-RKTDGFKNVEGFLIQHGTGDDNVHFQNSA 688  
DB 660 EYDSVYTERYMGFLTPEDNLDHYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717  
QY 689 ALVDLLMGDGVSPKHLHSQWFTSDHGISYHGGGVFLYKQLARKUYQ 735  
DB 718 QISKALVDVGD---FQAMWYTDHGIASSTAHQHIYTHMSHFQK 761

RESULT 14  
US-09-265-606-3  
; Sequence 3, Application US/09265606  
; Patent No. 6846910  
; GENERAL INFORMATION:  
; APPLICANT: Zimmermann, Rainer; Park, John E.;  
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.  
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
; TITLE OF INVENTION: ALPHA, AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felle & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/265,606  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/619,280  
; FILING DATE: 18-MARCH-1996  
; APPLICATION NUMBER: 08/230,491  
; FILING DATE: 20-APRIL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 6846910man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5330.1  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 766 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-09-265-606-3

Query Match 24.5%; Score 987.5; DB 2; Length 766;  
Best Local Similarity 32.6%; Pred. No. 9.3e-76;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TEGSGKRLTFNVTWKQAITPTSRVQWLSGAEDESYVYAAEDGSLTIENIVTNSRTLI 70  
DB 36 TADSRKTYLTLD-YLKNYRLKLYSLRWIS---DHEYLKQENNLVFNFAEYGNSSVFLE 91  
QY 71 PADKIPTGKEAFNYWIHPDLSSVWASNHTKQYRHSFFADYVQVDESLSKSVPLPDOOB- 129

```
Db 92 NSTDFEGHSHINDYSIPDQGFILLEYNYVQWRHSYASY---DIYDLNKRQLITERI 148
Qy 130 -GDIYAQWSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MPHGVPDVIYEEIL 185
Db 149 PNTQWTVWSPVGHKLAVWNNDIYVKEPNLPSYRITWTGKEDIINYNGITDWMYEEVF 208
Qy 186 GDRYALWSPDGEYLAIVLNFNETGVPTTYVOYYMNDQEIAPAYPWELKIRYPKVQNPT 245
Db 209 SAYSALWSPNGTFLAYAFQNDTEVPLIEYSFYSD---SLQYPKTVRVYPKAGAVNPT 265
Qy 246 VTLSLLNI--ASKEVKQAPIDAPESDIIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 266 VKFFVWNTDSSVNTNATSIQITAPASMLGDHYLDCVWTWATQER--ISLQWLRRIQNY 323
Qy 298 -QKVAVADTASNKATVISDRD----GTDGMDLNLMSKYGIPKPS-----DKDAYIID 346
Db 324 VMDICDYDESSGRWNCIVARQHIEMSTGM-----VGRFPSPHFTLDGNSFYKI 374
Qy 347 ISDHSGWAHLVLPVSGGEPILTKGDWEVTSILSDOERQLVYLTQHHSTE--RHLY 404
Db 375 ISNEEGYRHICYFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 432
Qy 405 SVSYSTFA-VTPL---VDDTVAAYWSASFNSANSYVILTYGGPDVPYQELVYTNSTKPLR 460
Db 433 KIQLSDYTKVTCSELNPERCQYYSVSFSKEAKYQLRCSGGPLPLYTLHSSVNDKGLR 492
Qy 461 TITDNA---KVLEQIK-----DYALPNITYFELPLPSGETLNMVQRLPGFSPDKKYP 511
Db 493 VLEDSALDKMLQNVQMPKSLDFIILNETKFWYQ-----ILPHFDKSKKYP 542
Qy 512 LFTPYGGGAQEVTKRQWALNFKAYVASDSLELYVTWVNRGTGFGKGRKFRSAVTQLG 571
Db 543 LLDVYAGPCSKADTVFR-LNWAYTLASTENI--IVASFDRGSGYQGDKIMHAINRRLG 599
Qy 572 LLEAEDQIYAAQQAANIPIWIDADHIGWGSFGYLTSKVLEKDSGAFLLGVIITAPYSDW 631
Db 600 TFEVEDQIEAARQFSKMGFVDNKRKIAIWGWSYGGYVTSWVLGSGGVKFCGIAVAPSRW 659
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSA 688
Db 660 EYDVSVYTERYMGUPTPEDNLDRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
Qy 689 ALVDLLMGDGVSPKHLQWFTSDSHGISYHGGGVFLYKQLARKLYQ 735
Db 718 QISKALVDVGD---FQAMWYTDHGHGASSTAHHQIYTHMSHFQ 761
```

```
RESULT 15
US-09-949-016-10450
; Sequence 10450, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10450
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10450
```

Query Match 24.5%; Score 987.5; DB 2; Length 775;

Best Local Similarity 32.6%; Pred. No. 9.5e-76;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

```
Qy 11 TGBGSKRLTNETVVKQAIPTPSRQVQWLSGABDSGVVYAAEDGSLTIENIVTNESSTLI 70
Db 45 TADSRKTYTLTD-YLKNTRYLKLKSLRWIS--DHELYLKQENNILVFNAEYGNSSVFLE 100
Qy 71 PADKIPGKEAFNYWHPDLSSVLWASNHTKQYRHSFFADYYVQDVESLSKSVLPDQOE- 129
Db 101 NSTDFEGHSHINDYSIPDQGFILLEYNYVQWRHSYASY---DIYDLNKRQLITERI 157
Qy 130 -GDIYAQWSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MPHGVPDVIYEEIL 185
Db 158 PNTQWTVWSPVGHKLAVWNNDIYVKEPNLPSYRITWTGKEDIINYNGITDWMYEEVF 217
Qy 186 GDRYALWSPDGEYLAIVLNFNETGVPTTYVOYYMNDQEIAPAYPWELKIRYPKVQNPT 245
Db 218 SAYSALWSPNGTFLAYAFQNDTEVPLIEYSFYSD---SLQYPKTVRVYPKAGAVNPT 274
Qy 246 VTLSLLNI--ASKEVKQAPIDAPESDIIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 275 VKFFVWNTDSSVNTNATSIQITAPASMLGDHYLDCVWTWATQER--ISLQWLRRIQNY 332
Qy 298 -QKVAVADTASNKATVISDRD----GTDGMDLNLMSKYGIPKPS-----DKDAYIID 346
Db 333 VMDICDYDESSGRWNCIVARQHIEMSTGM-----VGRFPSPHFTLDGNSFYKI 383
Qy 347 ISDHSGWAHLVLPVSGGEPILTKGDWEVTSILSDOERQLVYLTQHHSTE--RHLY 404
Db 384 ISNEEGYRHICYFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 441
Qy 405 SVSYSTFA-VTPL---VDDTVAAYWSASFNSANSYVILTYGGPDVPYQELVYTNSTKPLR 460
Db 442 KIQLSDYTKVTCSELNPERCQYYSVSFSKEAKYQLRCSGGPLPLYTLHSSVNDKGLR 501
Qy 461 TITDNA---KVLEQIK-----DYALPNITYFELPLPSGETLNMVQRLPGFSPDKKYP 511
Db 502 VLEDSALDKMLQNVQMPKSLDFIILNETKFWYQ-----ILPHFDKSKKYP 551
Qy 512 LFTPYGGGAQEVTKRQWALNFKAYVASDSLELYVTWVNRGTGFGKGRKFRSAVTQLG 571
Db 552 LLDVYAGPCSKADTVFR-LNWAYTLASTENI--IVASFDRGSGYQGDKIMHAINRRLG 608
Qy 572 LLEAEDQIYAAQQAANIPIWIDADHIGWGSFGYLTSKVLEKDSGAFLLGVIITAPYSDW 631
Db 609 TFEVEDQIEAARQFSKMGFVDNKRKIAIWGWSYGGYVTSWVLGSGGVKFCGIAVAPSRW 668
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSA 688
Db 669 EYDVSVYTERYMGUPTPEDNLDRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 726
Qy 689 ALVDLLMGDGVSPKHLQWFTSDSHGISYHGGGVFLYKQLARKLYQ 735
Db 727 QISKALVDVGD---FQAMWYTDHGHGASSTAHHQIYTHMSHFQ 770
```

Search completed: January 9, 2006, 03:21:31  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 03:20:31 ; Search time 76 Seconds  
(without alignments)  
4150.804 Million cell updates/sec

Title: US-10-723-807-2\_COPY\_17\_771

Perfect score: 4033

Sequence: 1 LDVPRKPAPTEGSGKRLTF.....EKNRQTQVLHMQHTKDLLE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	4033	100.0	771	4	US-10-723-807-2
2	2914	72.3	764	4	Sequence 2, Appli
3	2610.5	64.7	775	5	Sequence 12802, A
4	1269	31.5	790	4	Sequence 35, Appl
5	1162	28.8	793	4	Sequence 3803, Ap
6	1118	27.7	634	4	Sequence 2243, Ap
7	1054	26.1	818	4	Sequence 13304, A
8	1054	26.1	818	4	Sequence 3, Appli
9	1054	26.1	818	4	Sequence 3, Appli
10	1054	26.1	818	4	Sequence 3, Appli
11	1054	26.1	818	4	Sequence 22029, A
12	1054	26.1	818	6	Sequence 3, Appli
13	1054	26.1	818	6	Sequence 3, Appli
14	1054	26.1	818	6	Sequence 3, Appli
15	1054	26.1	818	6	Sequence 3, Appli
16	1054	26.1	818	6	Sequence 3, Appli
17	1020	25.3	504	4	Sequence 863, App
18	987.5	24.5	766	3	Sequence 3, Appli
19	987.5	24.5	766	3	Sequence 1, Appli
20	987.5	24.5	766	4	Sequence 590, App
21	987.5	24.5	766	5	Sequence 18, Appl
22	987.5	24.5	766	5	Sequence 565, App
23	982	24.3	760	5	Sequence 4171, Ap
24	981.5	24.3	766	4	Sequence 6, Appli
25	981.5	24.3	766	4	Sequence 7, Appli
26	981.5	24.3	766	4	Sequence 6, Appli
27	981.5	24.3	766	4	Sequence 922, App

ALIGNMENTS

RESULT 1

US-10-723-807-2  
; Sequence 2, Application US/10723807  
; Publication No. US20040171104A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Blinkovsky  
; APPLICANT: Kimberly Brown  
; APPLICANT: Michael W. Rey  
; APPLICANT: Alan Klotz  
; APPLICANT: Tony Byun  
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl  
; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same  
; FILE REFERENCE: 5254.200-US  
; CURRENT APPLICATION NUMBER: US/10/723,807  
; CURRENT FILING DATE: 2003-11-25  
; PRIOR APPLICATION NUMBER: US/09/079,592  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 08/857,884  
; PRIOR FILING DATE: 1997-05-16  
; PRIOR APPLICATION NUMBER: 60/062,892  
; PRIOR FILING DATE: 1997-10-20  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 771  
; TYPE: PRT  
; ORGANISM: Aspergillus oryzae  
US-10-723-807-2

Query Match	100.0%;	Score	4033;	DB	4;	Length	771;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	755;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	LDVPRKPAPTEGSGKRLTFNETVVKQAITPTSRVSQVLSGAEDGSGYVVAEDGSLITEN	60				
Db	17	LDVPRKPAPTEGSGKRLTFNETVVKQAITPTSRVSQVLSGAEDGSGYVVAEDGSLITEN	76				
QY	61	IVTNESTRLLIPADKIPTGKEAFNWIHPDLSSVLWASNHTKQYRHSFFADYVYQDVESLK	120				
Db	77	IVTNESTRLLIPADKIPTGKEAFNWIHPDLSSVLWASNHTKQYRHSFFADYVYQDVESLK	136				
QY	121	SVPLMPDQEGDIQYQAQWSPVGNITIAFVRENDLYVWDNGTVTRITDDGGDPMFHGVDPDWLY	180				
Db	137	SVPLMPDQEGDIQYQAQWSPVGNITIAFVRENDLYVWDNGTVTRITDDGGDPMFHGVDPDWLY	196				
QY	181	EEIIIGDRYALWFSFDPGEVLAFLSNFTGVPTTYQYYMNDNOEIPAYPWELKIRYPKVS	240				
Db	197	EEIIIGDRYALWFSFDPGEVLAFLSNFTGVPTTYQYYMNDNOEIPAYPWELKIRYPKVS	256				
QY	241	QTNPTVTLSLLNIASKEVKQAPIDAFESTDLIIGEVAWLTDTHTTVAAKFNRVQDQKV	300				

Db	257	QTNPTVTLLSLNIASKEYQKAPIDAFESTDIIIGEVAWLTDTHHTVAAKAFNRVQDQVK	316
Qy	301	VAVDTASNKATVISDRDGTGDLNLLSMKYIGIPKSDKDAYVIDISDHSGWAHLYLFP	360
Db	317	VAVDTASNKATVISDRDGTGDLNLLSMKYIGIPKSDKDAYVIDISDHSGWAHLYLFP	376
Qy	361	VSGGEPILPTKGDWEVTSILSIDQERQLVYVLTSTQHSTERHLYSVSYSTFAVTPLVDDT	420
Db	377	VSGGEPILPTKGDWEVTSILSIDQERQLVYVLTSTQHSTERHLYSVSYSTFAVTPLVDDT	436
Qy	421	VAAYSASFANSGYIITYGGPDVPYQELYYTNTSTKPLRTITDNAKVLEQIKDYALPNI	480
Db	437	VAAYSASFANSGYIITYGGPDVPYQELYYTNTSTKPLRTITDNAKVLEQIKDYALPNI	496
Qy	481	TYFELPLPSGETLNVQRLPQGFSPDKKYPILFPTYGPGAQEVTKRWQALNFKAYVSD	540
Db	497	TYFELPLPSGETLNVQRLPQGFSPDKKYPILFPTYGPGAQEVTKRWQALNFKAYVSD	556
Qy	541	SELEYVTVTVDNRGTFGRKFRSAVTSQLGLLEAEQIYAAQQAANIPWIDADHIGIWG	600
Db	557	SELEYVTVTVDNRGTFGRKFRSAVTSQLGLLEAEQIYAAQQAANIPWIDADHIGIWG	616
Qy	601	WSFGGVLTSKVLKESGAFITGVITAPVSDNRFPYDSMYTRYMKTLSITNEEGYETSVAVK	660
Db	617	WSFGGVLTSKVLKESGAFITGVITAPVSDNRFPYDSMYTRYMKTLSITNEEGYETSVAVK	676
Qy	661	TGQFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHSQWFTDSDHGISYHG	720
Db	677	TGQFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHSQWFTDSDHGISYHG	736
Qy	721	GGVFLYQLARKLYQEKNRQTQVLMHQTXXDLEE	755
Db	737	GGVFLYQLARKLYQEKNRQTQVLMHQTXXDLEE	771
RESULT 2			
US-10-369-493-12802			
; Sequence 12802, Application US/10369493			
; Publication No. US20030233675A1			
; GENERAL INFORMATION:			
; APPLICANT: Cao, Yongwei			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Chen, Xianfeng			
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION			
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
; FILE REFERENCE: 38-10(52052)B			
; CURRENT APPLICATION NUMBER: US/10/369,493			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US 60/360,039			
; PRIOR FILING DATE: 2002-02-21			
; NUMBER OF SEQ ID NOS: 47374			
; SEQ ID NO 12802			
; LENGTH: 764			
; TYPE: PRT			
; ORGANISM: Aspergillus nidulans			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1)..(764)			
; OTHER INFORMATION: unsure at all Xaa locations			
US-10-369-493-12802			

[illegible]

```
; OTHER INFORMATION: wherein Xaa can be any amino acid
US-10-926-188A-35

Query Match      64.7%; Score 2610.5; DB 5; Length 775;
Best Local Similarity 61.6%; Pred. No. 3.2e-203;
Matches 463; Conservative 134; Mismatches 148; Indels 7; Gaps 4;

QY 4 PRKPHPTGEGSKRLTFNETVVKQAITPTSRVQWLSGAEDGSGYVYAAEDSLTIENTVT 63
DB 19 PREPRSPPTGGNKLTYKCVPRATISPRSTSLAMINSEEDGRYISQSDDGALIIQNIVT 78

QY 64 NESRTLIPADKIPTCKEAFNYWIHPDLSVLWASNHTKQYRHSFPADYVQDVESLKVSP 123
DB 79 NTKTLVAADKVPKG--YYDYFKPDLSAVLWATNTYQYRHSYFANYFILDIKKGLTLP 136

QY 124 LMPDQEGDIOYQWSPVGNITIAFVRENDLYVNDGTVTRITDDGGDMFHGVPDWIYBEE 183
DB 137 LAQDAQGDIQYQWSPMNSIAVVRXNDLYIWNNGKTKRITENGCPDIFNGVPDMDYBEE 196

QY 184 ILGDRVALWFSPDGEYLAISFNETGVPTTYVQYMDNQEIAPAYPWELKIRYKPVQSTN 243
DB 197 IGDRLPALWFSPDGEYLAISFNETGVPTTYVQYMDNQEIAPAYPWELKIRYKPVQSTN 256

QY 244 PTVTLSLLNIASKYKQAPIDAFESTDLIIIGVAMLTDTHTTAAKAFNRVDDQOKVAV 303
DB 257 PTVQPHLLNIASSQETTPVTAAPPENDLVIGEVAWLSGGHDSVAVRAFNVRQDREKIVSV 316

QY 304 DTASKATVISDRDGTGDLNLLSMKYIGPIKPSDKDAYIIDISHSQWHLXLYLPVSG 363
DB 317 KYESKESKIRERDGTGDLNLLSMKYIGPIKPSDKDAYIIDISHSQWHLXLYLPVSG 373

QY 364 GPIPLTKGDWEVTSILSDORQOLVYVLTSTQHHSTERHLYSVSYSTFPAVTPLVDDTVAA 423
DB 374 GREIALTKGEWEVALLKVDTKKLLYFTSTKYHSTRHVSYSVDTKVTMTPLVNDKEAA 433

QY 424 YMSAFSANSYGYIITYGDPVPYQELVTT--NSTKPLRTITDPAKLEQIKDYALPNITY 482
DB 434 YITASFSAGGYIITYGDPVPYQELVTT--NSTKPLRTITDPAKLEQIKDYALPNITY 493

QY 483 FELPLPSGTEGLNVQRLPGFSPDKYPIPLTPYGGPGAQEVTKRQALNFKAYVASDSSE 542
DB 494 FBIKLPSGTEGLNVQRLPGFSPDKYPIPLTPYGGPGAQEVTKRQALNFKAYVASDSSE 553

QY 543 LEVVTWTVNDRGTGPKRFRSAVTRQLGLAEADQIYAAQQAANI--PWIDADHIGIWS 602
DB 554 LEVVTWTVNDRGTGPKRFRSAVTRQLGLAEADQIYAAQQAANI--PWIDADHIGIWS 613

QY 603 FGQYLTSKVLEKDSGAFTLGIVITAPVSDWRPFYDSMYTERYMKTLSTNEEGYETSARVKT 662
DB 614 YGGFLTAKTLETDGVTFTGISTAPVSDFRLYDSMYTERYMKTLSTNEEGYETSARVKT 673

QY 663 GPRQVEGGLIQTGTDGDDNVHFNQNSAALVDLLMGDGVSPKLSQWFTSDHGISVHGG 722
DB 674 GPKNLKGHYLIQHTGTDGDDNVHFNQNSAALVDLLMGDGVSPKLSQWFTSDHGISVHGG 733

QY 723 VELYKQARKLYOEKNRQIQ--VLMHQTWKOL 753
DB 734 TYQYKQLSKMVDQKRRPESPPMHQWSKRVL 765

RESULT 4
US-10-369-493-3803
; Sequence 3803, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3803
; LENGTH: 790
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3803

Query Match      31.5%; Score 1269; DB 4; Length 790;
Best Local Similarity 38.7%; Pred. No. 5.8e-94;
Matches 296; Conservative 132; Mismatches 271; Indels 66; Gaps 29;

QY 7 PHAPTGECS-KRLTFNETVVKQAITPTSRVQWLSG--AEDG-----SYVYAAD 53
DB 34 POATVTHSGSKVTLQDQ-VLNNWRPRAKSHSISWIAVNGEDGLLEKEGANKDYLVEDV 92

QY 54 GSLTIENIVTNESRTLIPADKIPTCKEAFN--YWIHP-----DLSSVLWASNHTKQYR 104
DB 93 RAQNPSSEVSEASKSTLI-KDKL---FEFANKTYM--PTVTVPSTRDLKKVLLATDVQNNR 146

QY 105 HSFPADYVQDVESLKSVPMLP--DQEGDIOYQWSPVGNITIAFVRENDLYV--WDNGTVT 161
DB 147 HSYAVVYVIFDVETQQAEPFLVYDADARLQLASWSPTSDAIVYTRDNNFLKRLSDKLV 206

QY 162 RTDDGGDPMFHGVDPDWIYBEEILGDRVALWFSPDGEYLAISFNETGVPTTYVQYTM-- 219
DB 207 QITRQGSADVFNQVDPDWIYBEEVVLASGVATWWSGDNVAVFRTNETGVPEPIQYFVSR 266

QY 220 --DNQBIAPA---YPWELKIRYKPVQSTNPTVTLSLLNIASKYKQAPIDA--FESTDLII 274
DB 267 PSGEPEKPEENYPEVRQIKYKPKAGAHNPIVDLKFDVVKRGDVFSDISGRFADDDRLIT 326

QY 275 EVAMLTDTHTTVAAKAFNRVQDQKVAVDATSN-----KATVISDRDGTGDLNLLSM 329
DB 327 EVIWM---AGKQVLIKETNRVSDMRVLDVDSRGTGKAVRTVVDVNDIDG--GMFEISHKT 381

QY 330 KYIGPIKPSD--KDAVYIIDISHSQWHL--YLFVPSGGGEPIPLTKGDWEVTSILS-IDOE 385
DB 382 KEI--PADFANGRPDGGYVDVTIHHNGDHLAYFTPLDNPIMPILTSGDYEVDAPSVDLQ 440

QY 386 ROLVYLTSTQHHSTERHLYSVSYSTFPAVTPLVDDTVAAAYWASFSANSYGYIITYGDPDV 445
DB 441 RNLVYFVSTKESSIQRHVYQVKLTGEDMTPTVDTSKGYIATSFSTGAGYALVSYQGPNI 500

QY 446 PYQELVTT--NSTKPLRTITDPAKLEQIKDYALPNITYFELPLPSGETLVNQRLPGCF 503
DB 501 PMQKVIPTSPNDPKYEHVVEENKDLAEAKGHELPIINIYGTINV-DGVELNIERRPPHF 559

QY 504 SPDKKYPIILFTPYGGPGAQEVTKRQALNFKAYVASDSSELYVVTWTVNDRGTGPKRFR 563
DB 560 DNKKYPIVLVFOQYSGVPSQTVKKTFF--AVDFQSFVAAG--LGVICVTVDRGRTGFGIKNR 616

QY 564 SAVTRQLGLAEADQIYAAQQAANI--PWIDADHIGIWSFGGYLTSKVLEKDSG--AFTLG 622
DB 617 VLIIRGNLGTWESHQDQIAAKHWAQKDYIDEDRLAIWGSYGGVYMTLKTLEQDAGQFKYG 676

QY 623 VITAPVSDWRPFYDSMYTERYMKTLSTNEEGYETSARVKTGDF--KNVEGGLIQTGTDGDDN 681
DB 677 MAVAPVTDWRPFYDSMYTERYMKTPQTNPEGYESAATVNTVNTALSQNR--FLMHGVADDN 734

QY 682 VHFQNSAALVDLLMGDGVSPKLSQWFTSDHGISVHGGVPLY 726
DB 735 VHMQNSLTLLDAL--DQRSVENYDQVFPDSDHGIYFFHNRILVF 777

RESULT 5
US-10-369-493-2243
; Sequence 2243, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
```

/ APPLICANT: Cao, Yongwei  
 /  
 / APPLICANT: Hinkle, Gregory J.  
 /  
 / APPLICANT: Slater, Steven C.  
 /  
 / APPLICANT: Goldman, Barry S.  
 /  
 / APPLICANT: Chen, Xianfeng  
 /  
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 /  
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 /  
 / FILE REFERENCE: 38-10(52052)B  
 /  
 / CURRENT APPLICATION NUMBER: US/10/369,493  
 /  
 / CURRENT FILING DATE: 2003-02-28  
 /  
 / PRIOR APPLICATION NUMBER: US 60/360,039  
 /  
 / PRIOR FILING DATE: 2002-02-21  
 /  
 / NUMBER OF SEQ ID NOS: 47374  
 /  
 / SEQ ID NO 2243  
 /  
 / LENGTH: 793  
 /  
 / TYPE: PRT  
 /  
 / ORGANISM: Schizosaccharomyces pombe  
 /  
 / US-10-369-493-2243

Query Match	28.8%	Score 1162;	DB 4;	Length 793;
Best Local Similarity	36.1%	Pred. No. 3e-85;		
Matches	272;	Conservative 142;	Mismatches 294;	Indels 46; Gaps 23;

  

Qy	22	ETVVKQAITPSRSVQWL--SGAEDGSVVAAEAGSLTIENIVNESRTLIP-ADKPIPTG	78
Db	58	DDIVLQKYKPSYKQVNWIDSGQLKD---TFVLVTKYGDGLINIQDPYNLNKLTFESVSDLVNIG	114
Qy	79	--KEAFNWHTPDLSSVLWASNNHTKQYRHSFPADYYVQDVESLKSVPIM--PDQEGDIQY	134
Db	115	IQLDYDSYSISFDAKYVLVSVNKSQRWHSFSAQYYLYNTET-KDVNMLGQDNEHWITSL	173
Qy	135	AQMSPVGNITAFVRENLDLYWDN-GTVTRITDDGGPMFHGVPDWIYEEELGRDYALWF	193
Db	174	AEMSPTGHQLSFVYNNDLYVRKNDGNVQRLTYDGTVDVFENGLTDWIYEEELVSPSTIWM	233
Qy	194	SPDGEYLAYLSFNETGVTPTTYQYVMDNOETAPAYP-----WELKIRPKVQSOTNPTVTL	248
Db	234	SPDSDKIAFLKLNSEIPTY--HYPLTYAEGLDPLPEFDYNKQMAIKYPKPGNPNPSVSL	291
Qy	249	SLNLIASKEVQAIPIDAPES-TDLIIIGEVAWLTDHTTVAAKAFNRVQDQKVAVDITAS	307
Db	292	FVADLNSNASSNFSLWNEPLAEVPVQNVLWNT--SSVLVQFTNENSTCITARLLDTEL	349
Qy	308	NKATVISDRDGTGDLNDLLSMKYIGPIKPS-----DKDAYIIDISDHSGWAHLYLFPVS	362
Db	350	KSIHTVKTCELEGWEYEQQSAKMF-PLNNSLVWENWSDGYF-DIALDDYNHLAFTPFN	407
Qy	363	GGEPIPLTKGWEVT-SILSDQERQLVYILSTQHHSTERHLYSVSYSTFATVPLVDD-T	420
Db	408	GSSPIYLTSGAWDVTGGPIHTDGGFGNVYFLATLKDSTERHLYVYSLDTLBIYGITDNGE	467
Qy	421	VAAWKSASFNSANGYIILTYGGPDVPYQELYTNTSTKPLRTITONAKVLEQIKQVAPENI	480
Db	468	DEGYSTSFSPFGFYVLVNHGPDVPWQELKSTKDQCYCLETNRSRUKQQLSITLPSV	527
Qy	481	TYFELPLPSGDTLNVQRLPPGSPDKKYPILFPTPYGGPQAQEVTKRQWALNFKAYASD	540
Db	528	EYKLTLP-NDTTFNMEERRPNFDVNKKYPVLFFAYGGPGSQQVAKLFR-VDFQAYLASH	585
Qy	541	SELEYVTWTDVNRGTGFKGRKFRSAVTROGLLEABDQIYAAQQAANTIPWIDAHIGTWG	600
Db	586	PDFFIVVTLDRGTGFGNAGAFRYSVRHLGEWESYDQQAQKFWADLPFDVDENHVGIWG	645
Qy	601	WSFGYLTSKVLKXDSGAFTLGVIAPYSDRRFYDSMYTERMKTLSTNESGYETSAVRK	660
Db	646	WSYGGYLLTKLTLETQD-VFYSYGMAVAPYDMLRYLDSVYTERYMDLPQYKNGKYGKNSQIHD	704
Qy	661	TGDFKNWEGGFLIOHGTGDDNVHFONSAAVLDMGMGDGVSPEKLHS---QWFTDSDHGIS	717
Db	705	YEKFKQLK-RFVFAHGTGDDNVHFGHSHML-----DGLNLANCYNIDMAVFPDPSAHSIS	758
Qy	718	YHGGGVFLYKQLARKLYQOEKNR---QTQVLMHQW	748

```

Db      759 YHNASLSIYHRLSEWIGDALGRIDPSTGVRQHRW 792

RESULT 6
US-10-369-493-13304
; Sequence 13304, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13304
; LENGTH: 634
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(634)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13304

```

Query Match	27.7%	Score 1118;	DB 4;	Length 634;
Best Local Similarity	38.1%;	Pred. No. 8.4e-82;		
Matches 246;	Conservative 105;	Mismatches 237;	Indels 58;	Gaps 17;
QY	129	EGDIQYQWSPVGNITIAFVRENDLYV--WDMGTVTRITDDGGPDMFHGVPDWMYEEIEILG	186	
DB	2	DGRVQLALPLNSDAIAFVDDNNLYLRKLSSEVVPIKDGQEQLFYGVDPDWYEEVEFS	61	
QY	187	DRYALWSPDGEYLAFLSFNETGPTTVYQYMDN-----OETAPAYPWLKRYPKVS	240	
DB	62	GNSVTWMSDGSQAFIRTINESAVPEFPVQYFLSRPGKKPQGLNYPFVREIKYIPKAG	121	
QY	241	QTNPTVTLSLNIASKEVKQAPI-DAPESDILLIIEVAWLTDTHTTAAKAFNRVQDOOK	299	
DB	122	APNFVNQLQFVDEQGEVFSVDTFDDDDDLIIIEVIA--AKGKLVRTTNRESDILK	179	
QY	300	VVAVDITASNKATVSDRDGTD--GWLNDLLSMKYIGPIKPSDKDA-----YYIDISDS	351	
DB	180	VFLVDTSRESKLIRIQDIAELDGGWBEPTQSVRFI----PADPDKGRFPFGDLYTVVHE	235	
QY	352	GWAHL-YLFPVSGGCEPLTKGDMEVTSI-LSIQEROLVYVLSLTOHSTHERHLYSVSYS	409	
DB	236	GYDHLVYPTLDNPEPIMLTSGEWEVDAPTAVDLTRLGLVFIATKEAPTEHLRYRLD	295	
QY	410	TFAVTPLVDDTVAAYSWASFSANGSYIILTYGGPDVPYQELVYT--NSTKPLRTITDNAK	467	
DB	296	GSDLTFLDTSQPGYYSVFSGAGYALLSYQGPSIPWQSIISTEGEKTTTLRIIEDNTD	355	
QY	468	VLEQIKDYALNIIYFELPLPSGETINWQRLPGCFSPDKVKYPILFTPYGGPGAQEVTKR	527	
DB	356	LSKLVAQYALPTENYQNITI-DGFTLQVVERRPHFNFPARKYPLVFLHYLGGPGSTVDNR	414	
QY	528	WQALNFKAIVASDSLELYVTWVDNRGTGFKGRKFRSAVTRQLGLLEAEDQIYAAQQAAN	587	
DB	415	FN-VDFAQSYVA--ASLGIYIVTVDGRGTGFGIARAARCIIRGNIGHYEAIQDIATAKNNAQ	471	
QY	588	IPWIDADHIGWGSFGGLYTSKVLKDSG-AFTLGVITAPVSDWRFYD-----	635	
DB	472	KPYVDESMAIWGWSYGGFMTLKTLQDAGETFYQGMVAVPTDWRFTKGLKPTHFLLG	531	
QY	636	-----SMYTERYMKTLSTNEEGYETSVAVRKTDGPFKNVGGFLIQHGTGDDNVH	683	

```
Db 532 PCIVNKPRAQXISVYTERYMTPOHNPCTGYDNTSISDMAALHN-NVRFVLVHGASDDNVH 590
Qy 684 FQNSAALVDLLMGDGVSPKLSHQWFTDSDHGIGSYHGGGVFLYKOL 729
Db 591 IONTLTLDKGL--DLASQVNDVHFYPSDSDHSHIFFHNAHTWVIERM 634

RESULT 7
US-10-401-437-3
; Sequence 3, Application US/10401437
; Publication No. US20030158070A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,437
; CURRENT FILING DATE: 2003-03-26
; PRIOR FILING DATE: 2003-03-26
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-401-437-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNFTVVKQAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYTNDGKLVKVSFS-VVRNNTFPQPKYHBLQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLPADKIPGKEAFNWIH-----PDLSSVLW 95
Db 107 TFMND-SYVVKSVYDSDSYNSVLLEGKT-----FIHQNLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQYRHSFADYVYQDVESLKSVPMLPQEGDIQYAAQSPVGNITAFVRENDLYW 155
Db 157 RINSQVNRHSITFGSYFYVD----KSSSFEEIGNEVALAIWSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPWIYEEILGDRYALWFPDGEYLAFLSNFTGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAAWSPDGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDQNEIAPAPWELKIRYKVSQTNPTVTLSLNLNIAKESVKEVQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMERSIKYKSGTNPFAELWVYSMDKGTGFHPRISGNKKGSL 330
Qy 272 IIGEVAVLTDTHTTAAKAFNRVQDQKQVAVDTSNKAATVISDRDGTGDLNLLSMKY 331
Db 331 LITEVTVNGN--VLVKTDRSSDLITVFLDITAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISDHSGWAHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDERQL 388
Db 389 IPANETFDPRHNGYVDILPIGGYNHLAYFENSNSHYKTLTEGKWEVNGPLAFDSMENR 448
Qy 389 VYVLSQHHSTERHLYSVSY-STFAVTPLVDDTVAAVWSASFSANGSYILTYGGPDVY 447
Db 449 LYPISRKSTSRHRYIIDLRSNIEIVTDSGDYDVYSSGRRFGLLYTKGKPVY 508
Qy 448 QELYTNTST-----KPLRTITONAKVLEQIKDYALPNITYFELPLPSGE----491
Db 509 QKIVDHSRKAECCKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLKGFQKD 568
Qy 492 -TLNVQRILPPGSP--DKKYPILFTPYGPGQAEVTKRWQALNFKAYVASDSELYVTW 548
Db 569 ILVNSYEILLPNDPDELTDLSHPYVPFFAYGPNQQVVKTF-SVGFNEVVA--SOLNAIVV 625
```

```
Qy 549 TVDNRGTGFKGRKFRSAVTRQIGLLEABEDQIYAAQAAANIPIWDADHIGIWCWSEGGYLT 608
Db 626 VVDGRGTGFKQDQFRSLVRDLGDEYARDQISAAISLYGSLTFVDPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVTAPVSDNRFDYSMTYTERYMKLTSTNEEGYETSAVRKTDGKXNV 667
Db 686 LKLTLEKQGRHFKYGMVAPVTDWRFDYSVYTERYMKHTPOENFDGYVSSVHNVTALAQA 745
Qy 668 EGGFLIQHTGDDNVHVFQNSAALVDLLMGDGVSPKLSHQWFTDSDHGIGSYHGGGVFLYK 727
Db 746 -NRFLMGTGDDNVHVFQNSLKFLLDLLLNGVENVYVHV--FPDSDHSIRYHNANVIVPD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 8
US-10-402-312-3
; Sequence 3, Application US/10402312
; Publication No. US20030166248A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/402,312
; CURRENT FILING DATE: 2003-03-26
; PRIOR FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-402-312-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNFTVVKQAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYTNDGKLVKVSFS-VVRNNTFPQPKYHBLQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLPADKIPGKEAFNWIH-----PDLSSVLW 95
Db 107 TFMND-SYVVKSVYDSDSYNSVLLEGKT-----FIHQNLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQYRHSFADYVYQDVESLKSVPMLPQEGDIQYAAQSPVGNITAFVRENDLYW 155
Db 157 RINSQVNRHSITFGSYFYVD----KSSSFEEIGNEVALAIWSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPWIYEEILGDRYALWFPDGEYLAFLSNFTGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAAWSPDGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDQNEIAPAPWELKIRYKVSQTNPTVTLSLNLNIAKESVKEVQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMERSIKYKSGTNPFAELWVYSMDKGTGFHPRISGNKKGSL 330
Qy 272 IIGEVAVLTDTHTTAAKAFNRVQDQKQVAVDTSNKAATVISDRDGTGDLNLLSMKY 331
Db 331 LITEVTVNGN--VLVKTDRSSDLITVFLDITAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISDHSGWAHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDERQL 388
Db 389 IPANETFDPRHNGYVDILPIGGYNHLAYFENSNSHYKTLTEGKWEVNGPLAFDSMENR 448
Qy 389 VYVLSQHHSTERHLYSVSY-STFAVTPLVDDTVAAVWSASFSANGSYILTYGGPDVY 447
```

```
Db 449 LYFISTRKSTERHVVYIDLRSPNEIIEVTDTSDDGVYDVVSFSSGRFPGLLTYKGPVPY 508
Qy 448 QELYTTNST-----KPLRTITDNAKVLQIKDYALPNITYFELPLPSGE---- 491
Db 509 QKIVDFHSRAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGFS--DKKYPILFTPYGGGAQEVTRKWOALNFKAYVASDSLELYVTW 548
Db 569 ILVNSYELPNDPDETILSDHYPPVFFAYGGPNSQOVVKTf--SVGFNEVVA--SQLNAIVV 625
Qy 549 TVDNRGTFGKGRKFRSAVTRQLGLEAEADQIYAAQQAANIPIWDADHIGIWMGWSFGVLT 608
Db 626 VVDRGRTGFGKQDFRSLVRDLRGDYEARQDQISAASLYGSLTFVDPQKISLFGWSYGGVLT 685
Qy 728 QL 729
Db 803 KL 804

RESULT 9
US-10-402-067-3
; Sequence 3, Application US/10402067
; Publication No. US20030175892A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; CURRENT APPLICATION NUMBER: US/10/402,067
; PRIOR FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-402-067-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTFTNETVVKQAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQEPNSNYTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTILIPADKIPGKEAFNYWIH-----POLSSVLW 95
Db 107 TFMND-SYVVKSVYDDSYNSVLLEGT-----FIHNGQLTVESITASPDLKRLLI 156
Qy 96 ASNHTKOYRHGFFADYVYQDVVSVLPSKVPMLPQDEGDIQYAOQSPVGNITAFVRENDLYW 155
Db 157 RTNSVQNRHSHSTFGSYFVVD---KSSSFEIENIGNEVALATWSPNSNDIAYVQDNNIYI 212
Qy 156 ---DNGVTTRITDDGPDHFCVPMWYEEELIGDRIYALWSPDCEYLAYLSFNETGYPT 212
Db 213 SAISKKTIRAVTNDGSSFLNFKPDMWYEEEFDDKAAMWSPTGDYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQETAPAPWELKIRYKVSQTNPTVTLSLNIATSKVEKQAPIDAFEST-DL 271
Db 273 FIIPYVQDEK--DIYPENRSIKYKSGTTPNPAELWVYSKMDGTSFPHRISGNKKGSL 330
```

```
Qy 272 IIGCEVAWLDTHTTTAAKAFNRVQDQKVVAVDTASNKATVVISDRDCTDGDWLDNLSMKY 331
Db 331 LITEVTVWNGN--VLVKTTDRSSDILTVFLDTIATKTSNVVRNNESSNGHWEIHTWTLF 388
Qy 332 IGPIKPSDDKAY-YIDISDHSWAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
Db 389 IPANETFDPRHNGYVDILPIGGYNHLAYFENSSSHYKTLTEGKWEVVGNGPLAFDSMENR 448
Qy 389 VYLSSTOHSTERHLYSVSY-STFAVTPLVDDTVAAYTWSASFANSYIILTYGPDVPY 447
Db 449 LYFISTRKSTERHVVYIDLRSPNEIIEVTDTSDDGVYDVVSFSSGRFPGLLTYKGPVPY 508
Qy 448 QELYTTNST-----KPLRTITDNAKVLQIKDYALPNITYFELPLPSGE---- 491
Db 509 QKIVDFHSRAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGFS--DKKYPILFTPYGGGAQEVTRKWOALNFKAYVASDSLELYVTW 548
Db 569 ILVNSYELPNDPDETILSDHYPPVFFAYGGPNSQOVVKTf--SVGFNEVVA--SQLNAIVV 625
Qy 549 TVDNRGTFGKGRKFRSAVTRQLGLEAEADQIYAAQQAANIPIWDADHIGIWMGWSFGVLT 608
Db 626 VVDRGRTGFGKQDFRSLVRDLRGDYEARQDQISAASLYGSLTFVDPQKISLFGWSYGGVLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDWRFYDSMYTERYMKTLSTNBERGYETSARVKTGDFKNV 667
Db 686 LKLTLEKDGGRHFKYGMVAPVTDWRFYDSVYTERYMTPOENFQYVESVHNVTALAAQ 745
Qy 668 EGGFLIOHGTGDDNVHFNQNSAALVDLLMGDGVSPKHLHSQWFTSDHGHSYHGGGVFLYK 727
Db 746 -NRELLMHGTGDDNVHFNQNSLKFLDLLDLNGVENYDVHV--FPDSHHSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 10
US-10-401-436-3
; Sequence 3, Application US/10401436
; Publication No. US20030180932A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,436
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-401-436-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTFTNETVVKQAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQEPNSNYTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTILIPADKIPGKEAFNYWIH-----POLSSVLW 95
Db 107 TFMND-SYVVKSVYDDSYNSVLLEGT-----FIHNGQLTVESITASPDLKRLLI 156
Qy 96 ASNHTKOYRHGFFADYVYQDVVSVLPSKVPMLPQDEGDIQYAOQSPVGNITAFVRENDLYW 155
Db 96 ASNHTKOYRHGFFADYVYQDVVSVLPSKVPMLPQDEGDIQYAOQSPVGNITAFVRENDLYW 155
```



Db 157 RTNSVQNRHSTFGSVFYVD-----KSSSFEBEIGNEVALAIWSPNSNDIAYVQDNNIYI 212  
 Qy 156 ---DNGTTRITDDGPDHFGVDPWIIYEEELGDRYALWSPDCEGYLAYLSFNETGVTPT 212  
 Db 213 SAISKTTIRAVTNDGSSFLFNGKPDWVYEEEFEDDKAAWMSPTGYLAFLKIDSEVGE 272  
 Qy 213 YTVQYMDNQEIAPAYPMELKIRYPKVQSTNPTVTLISLNIASKEVKQAPIDAFEST-DL 271  
 Db 273 FIIPYVQDEK--DIYPEMRSIKYPKSGTPNPHAEIWWYSMDKGTSPHPRISGNKKDGL 330  
 Qy 272 IIGEVAMLTDTHTTAAKAFNRVQOQKVAVDTASNKATVLSDRDGTGDLNLLSMKY 331  
 Db 331 LITEVTWVGNG--VLVKTDRSSDILTAVFLDITIAKTSNVVRNENSSNGMWEITHNTLF 388  
 Qy 332 IGPIPSDKDAY-YIDISDHSQWALHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388  
 Db 389 IPANETDPRPHNGYVDILPIGYNHLAYFENSNSHYKTLTEGKWEVNGPLAFDSMENR 448  
 Qy 389 YVYLSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSNGYVILTYGGPDVPT 447  
 Db 449 LFISTRKSSTERHVIYIDLRSPNEIIEVTDTSDEGVYDVSVSSGRRFGLLTYGPKVPY 508  
 Qy 448 QELYTNTST-----KPLRTITDNAKVLBOIKDYALPNITYFELPLPSGE---- 491  
 Db 509 QKIVDFHSRKABKCDKGNVLSKLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568  
 Qy 492 -TLNVMORLPPGFSP--DKKYPILFTPYGPGAGAEVTKRQWALNFKAYVASDSELEYVTW 548  
 Db 569 ILVNSYEILPNDFETLSHDYVFPFAYGPGNSQQVVKTF-SVGFNEVVA--SOLNAIVV 625  
 Qy 549 TVDNRGTGPKGRKFRSAVTRQGLLEAEQIIVAAQQAANIPIWDADHIGWMSFGGYLT 608  
 Db 626 VVDGRGTGFGQDFRSLVRDLGDEARQIISAALYGLSLTFVDPQKISLFGWSYGGYLT 685  
 Qy 609 SKVLEKDSGA-FTLGVITAPVSDMRFYDSMYTERYMKTLSTNEEGYETSVAVRTDGFKNV 667  
 Db 686 LKTEKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHHTPOENPDFGVSSEVSHVNTALAQA 745  
 Qy 668 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSDHGISHYGGVFLYK 727  
 Db 746 -NRFLMHGTGDDNVHFQNSKFLDLDDLNGVENYDVHV--FPDSDHSIRYHNAVIVFD 802  
 Qy 728 QL 729  
 Db 803 KL 804

RESULT 11  
 US-10-369-493-22029  
 ; Sequence 22029, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 22029  
 ; LENGTH: 818  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 US-10-369-493-22029  
 Query Match 26.1%; Score 1054; DB 4; Length 818;  
 Best Local Similarity 33.1%; Pred. No. 2e-76;

Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;  
 Qy 3 VPRKPHAP-----TGEESKRLTFNETVVKQAITPTSRSVQMLS-----GAEDGSYV 48  
 Db 48 IPHNSNTDYQEPNSNYTNDGKLKVSFS-VVRNTTFQPKYHELOMSDNKIESNDLGLV 106  
 Qy 49 YAAEDGSJTIENTVNESRTLIPADKIPTGKEAFNYWH-----PDLSSVLW 95  
 Db 107 TFMND-SYVAKSVYDDSYNSVLEGT-----FIHNGQNLTVESTASPDKRLLI 156  
 Qy 96 ASNHTQYVRHSPFADYVQDVESLAKSVPLMPDQEGDIOVAQWSPVGNITAFVRENLDLYW 155  
 Db 157 RTNSVQNRHSTFGSVFYVD-----KSSSFEBEIGNEVALAIWSPNSNDIAYVQDNNIYI 212  
 Qy 156 ---DNGTTRITDDGPDHFGVDPWIIYEEELGDRYALWSPDCEGYLAYLSFNETGVTPT 212  
 Db 213 SAISKTTIRAVTNDGSSFLFNGKPDWVYEEEFEDDKAAWMSPTGYLAFLKIDSEVGE 272  
 Qy 213 YTVQYMDNQEIAPAYPMELKIRYPKVQSTNPTVTLISLNIASKEVKQAPIDAFEST-DL 271  
 Db 273 FIIPYVQDEK--DIYPEMRSIKYPKSGTPNPHAEIWWYSMDKGTSPHPRISGNKKDGL 330  
 Qy 272 IIGEVAMLTDTHTTAAKAFNRVQOQKVAVDTASNKATVLSDRDGTGDLNLLSMKY 331  
 Db 331 LITEVTWVGNG--VLVKTDRSSDILTAVFLDITIAKTSNVVRNENSSNGMWEITHNTLF 388  
 Qy 332 IGPIPSDKDAY-YIDISDHSQWALHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388  
 Db 389 IPANETDPRPHNGYVDILPIGYNHLAYFENSNSHYKTLTEGKWEVNGPLAFDSMENR 448  
 Qy 389 YVYLSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSNGYVILTYGGPDVPT 447  
 Db 449 LFISTRKSSTERHVIYIDLRSPNEIIEVTDTSDEGVYDVSVSSGRRFGLLTYGPKVPY 508  
 Qy 448 QELYTNTST-----KPLRTITDNAKVLBOIKDYALPNITYFELPLPSGE---- 491  
 Db 509 QKIVDFHSRKABKCDKGNVLSKLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568  
 Qy 492 -TLNVMORLPPGFSP--DKKYPILFTPYGPGAGAEVTKRQWALNFKAYVASDSELEYVTW 548  
 Db 569 ILVNSYEILPNDFETLSHDYVFPFAYGPGNSQQVVKTF-SVGFNEVVA--SOLNAIVV 625  
 Qy 549 TVDNRGTGPKGRKFRSAVTRQGLLEAEQIIVAAQQAANIPIWDADHIGWMSFGGYLT 608  
 Db 626 VVDGRGTGFGQDFRSLVRDLGDEARQIISAALYGLSLTFVDPQKISLFGWSYGGYLT 685  
 Qy 609 SKVLEKDSGA-FTLGVITAPVSDMRFYDSMYTERYMKTLSTNEEGYETSVAVRTDGFKNV 667  
 Db 686 LKTEKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHHTPOENPDFGVSSEVSHVNTALAQA 745  
 Qy 668 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSDHGISHYGGVFLYK 727  
 Db 746 -NRFLMHGTGDDNVHFQNSKFLDLDDLNGVENYDVHV--FPDSDHSIRYHNAVIVFD 802  
 Qy 728 QL 729  
 Db 803 KL 804

RESULT 12  
 US-11-014-051-3  
 ; Sequence 3, Application US/11014051  
 ; Publication No. US20050095683A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Estell, David A.  
 ; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
 ; FILE REFERENCE: GC382-US  
 ; CURRENT APPLICATION NUMBER: US/11/014,051  
 ; CURRENT FILING DATE: 2004-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/US98/14647  
 ; PRIOR FILING DATE: 1998-07-14  
 ; PRIOR APPLICATION NUMBER: EP 97305237.7  
 ; PRIOR FILING DATE: 1998-07-15

```
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-11-014-051-3

Query Match      26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 IPHHSNTPDYQEPNSNTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 YAAEDGSLTIENITVNESRTLIPADKIPTGKEAFNWIH-----FIHNGQLTVESITASPDRLKLLI 156
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 TFMND-SYVVKSVYDDSYNSVLLGKT-----FIHNGQLTVESITASPDRLKLLI 156

Qy 96 ASNHTKQYRHSFFADYIVQDVESLSKSVLMPDQEGDIQYAQWSPVGNVTIAPVRENDLYVW 155
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
157 RTNSVQVNRHSTFGSYFYVD-----KSSSSPFEIGNEVALAIWSPNSNDIAVQDNNIYIY 212
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 ---DNGVTTRITDDGGPDMFHGVDWIYEBEILGDRYALWFSPPGEYLAVLSFNETGVPT 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 SAISKKTIRAVTNDGSSFLFNGKPDWYEEVEFDDKAAMWSPTGDIYLAFLKIDSEVGE 272
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 YTVQYVYVQDEK--DIYPEMRSIKYKPGTGNPHAEALWVYSKMGCTSFHPRISGNKKDGLS 330
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
272 IIGVAVMLTDHTTVAAKAFNRVQDQKVAADVATSNKATVISDRDGTGDLNLLSMKY 331
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
331 LITEVTVWGN--VLVKTTRDSSDILTFLIDTIAKTSNVVRNESSNGGWETHNTLTF 388
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 IGPIKPSDKDAY-YIDISDHSGMAHLVLPVSGGEP--PLTKGDWEVTS-ILSIDQEROL 388
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 IPANETDRPHNGVVDILPIGYNHAYFENSNSHYKTLTEGKWEVVGPLAFDSMENR 448
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 VYILSTQHHSTERHLYSVSY-STFAVTLVDVTVAAYWSAFSANGSYIILTYGGPDVPY 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
449 LYFISTRKSTERHVIYIDLRSPNEIIEVTDSEGVYDVVSFSSGRFRGLLTYGPKVPY 508
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
448 QELTYTNTST-----KPLRTITDNAKLVLEQIKDYALPNITYPELPLPSGE---- 491
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
509 QKIVDFHSRKAECDCGKGNVLGKSLYHLEKNEVLTKILEDYAVPRKS PRELNLGKDFGKD 568
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
492 -TLNMQRLLPPGSP--DKKYPILFTPYGGGAGAEVTKRWQALNFKAVASDSSELYVTW 548
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
569 ILVNSYEILPNDFEETLSHDYVPVFFAYGPGNSQQVKTFF-SVGNEVVA--SQLNAIVV 625
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
549 TVDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWADHIGIWSFGGYLT 608
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 VVDGRGTGFKQDQDFRSLVRDLRGDYEARQDI SAASLYGSLTFVDPDKLSLFGWSYGGYLT 685
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
609 SKVLEKDSGA-FTLGVITAPVSDMRFYDSMYTERYMKTLSTNEEGYETS SAVRKTDFGNV 667
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
686 LKTLKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHPTPQENFDFGVSSVHNVTALQA 745
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 EGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTSDSHGISYHGGGVFLYK 727
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
746 -NRFLLMHGTGDDNVHFQNSLKFLLDLLNGVENYDVHV--FPDSDSHSIRYHNANVIVFD 802
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
728 QL 729
803 KL 804

RESULT 13
US-11-013-991-3
; Sequence 3, Application US/11013991
; Publication No. US20050101001AI
```

```
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/013,991
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US 09/462,845
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-11-013-991-3
```

```
Query Match      26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 IPHHSNTPDYQEPNSNTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 YAAEDGSLTIENITVNESRTLIPADKIPTGKEAFNWIH-----FIHNGQLTVESITASPDRLKLLI 156
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 TFMND-SYVVKSVYDDSYNSVLLGKT-----FIHNGQLTVESITASPDRLKLLI 156

Qy 96 ASNHTKQYRHSFFADYIVQDVESLSKSVLMPDQEGDIQYAQWSPVGNVTIAPVRENDLYVW 155
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
157 RTNSVQVNRHSTFGSYFYVD-----KSSSSPFEIGNEVALAIWSPNSNDIAVQDNNIYIY 212
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 ---DNGVTTRITDDGGPDMFHGVDWIYEBEILGDRYALWFSPPGEYLAVLSFNETGVPT 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 SAISKKTIRAVTNDGSSFLFNGKPDWYEEVEFDDKAAMWSPTGDIYLAFLKIDSEVGE 272
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 YTVQYVYVQDEK--DIYPEMRSIKYKPGTGNPHAEALWVYSKMGCTSFHPRISGNKKDGLS 330
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
272 IIGVAVMLTDHTTVAAKAFNRVQDQKVAADVATSNKATVISDRDGTGDLNLLSMKY 331
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
331 LITEVTVWGN--VLVKTTRDSSDILTFLIDTIAKTSNVVRNESSNGGWETHNTLTF 388
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 IGPIKPSDKDAY-YIDISDHSGMAHLVLPVSGGEP--PLTKGDWEVTS-ILSIDQEROL 388
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 IPANETDRPHNGVVDILPIGYNHAYFENSNSHYKTLTEGKWEVVGPLAFDSMENR 448
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 VYILSTQHHSTERHLYSVSY-STFAVTLVDVTVAAYWSAFSANGSYIILTYGGPDVPY 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
449 LYFISTRKSTERHVIYIDLRSPNEIIEVTDSEGVYDVVSFSSGRFRGLLTYGPKVPY 508
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
448 QELTYTNTST-----KPLRTITDNAKLVLEQIKDYALPNITYPELPLPSGE---- 491
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
509 QKIVDFHSRKAECDCGKGNVLGKSLYHLEKNEVLTKILEDYAVPRKS PRELNLGKDFGKD 568
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
492 -TLNMQRLLPPGSP--DKKYPILFTPYGGGAGAEVTKRWQALNFKAVASDSSELYVTW 548
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
569 ILVNSYEILPNDFEETLSHDYVPVFFAYGPGNSQQVKTFF-SVGNEVVA--SQLNAIVV 625
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
549 TVDNRGTGFKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWADHIGIWSFGGYLT 608
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 VVDGRGTGFKQDQDFRSLVRDLRGDYEARQDI SAASLYGSLTFVDPDKLSLFGWSYGGYLT 685
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
609 SKVLEKDSGA-FTLGVITAPVSDMRFYDSMYTERYMKTLSTNEEGYETS SAVRKTDFGNV 667
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
686 LKTLKDGGRHFKYGMVAPVTDWRFYDSVYTERYMHPTPQENFDFGVSSVHNVTALQA 745
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 EGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTSDSHGISYHGGGVFLYK 727
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 EGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTSDSHGISYHGGGVFLYK 727
```

```
Db 746 -NRFLMHGTGDDNVHFNQSLKFLDLLDNGVENVYDVHV--FPDSHDSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 14
US-11-014-339-3
; Sequence 3, Application US/11014339
; Publication No. US20050142638A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/014,339
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US 09/462,845
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-11-014-339-3

Query Match 26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYNDGKLVSFVS-VVRNNTFPQKYHQLWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLPADKIPGKEAFNYIH-----FIHNGQLTVESITASPDLLKLLI 156
Db 107 TFMND-SYVVKSVYDSDSYNSVLEGGT-----FIHNGQLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQVHSPFADYVQDVESLKSVPMPQEGDIQYAAQSPVGNITAFVRENDLYW 155
Db 157 RTNSQVNRHSTFGSYFYVD-----KSSSFEEIGNEVALAIWSPNSNDIAYVDNNIYI 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPDWIYEEIILGDRYALWFGPDGEYLAFLSNFTGVP 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFPEDDKAAWWSPTGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPYWPWELKIRYKVSQTNPTVTLISLNIAASKEVKQAPIDAFEST-DL 271
Db 273 FIIPYVQDEK--DIYPEMRSIKYPKSGTPNPHAEALWVYSMDKGTGFPHRISGNKKGSL 330
Qy 272 IIGEVANLTDTHTTAAKAFNRVQDQKVAVDVASNKATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTTRDSSDLTLVFLDITIAKTSNVVRNESSNGGWETHNTLTF 388
Qy 332 IGPIKPSDKDAY-YIDISDHSQWHAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
Db 389 IPANETFDPHNGYVDILPIGGFNHLAYFENSGNSHYKTLTEKWEVNGPLAFDSMENR 448
Qy 389 VYVLSQTHSTHERHLSVSY-SFPAVTPLVDDTVAAAYWSASFANSYGYLYTGGPDVPY 447
Db 449 LFISTRKSTERHYIIDLRSPNEIIEVTDISEDGVYDVSFSGSRGFLGITYKGPVY 508
Qy 448 QELYTNTST-----KPLRTITONAKVLEQIKDYALPNITYIFELPLPSGE---- 491
Db 509 QKIVDPHSRKAECRKGNGVLGSLYHLEKNEVLTKILEDYAVPRKSFREINLGKDFGKD 568
Qy 492 -TLNVNQRLPPGESP--DKKYPILFTPYGGPQAQEVTKRQALNFKAYVASDSELEYVTW 548
```

```
Db 569 ILVNSYEILPNDPDETLSDHYPEFFFAYGSPNSQVWKTF-SVGNEVVA--SQLNAIVV 625
Qy 549 TVDNRGTCGKGRKERSAVTROGLLEAEADQIYAAQQAANIPIWDADHIGIWSFGGYLT 608
Db 626 VVDGRGTGFGKQDFRSLVRDLGDYEARDQISASLYGSLTFDVPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDMRFRYDSMYTERYMKTLSTNEBGEYSAVRKTGDFKNV 667
Db 686 LKTLKOGGRHFKYGMVAPVTDWRFYDSVYTERYMHTPQENFDGYVSVSSVHNVTALAQA 745
Qy 668 EGGFLIQHGTGDDNVHFNQSAALVDLLMGDGVSPKLSHSONFTDSDHGISVHGGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFNQSLKFLDLLDNGVENVYDVHV--FPDSHDSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 15
US-11-014-364-3
; Sequence 3, Application US/11014364
; Publication No. US20050214616A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/014,364
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US 09/462,845
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-11-014-364-3

Query Match 26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYNDGKLVSFVS-VVRNNTFPQKYHQLWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLPADKIPGKEAFNYIH-----FIHNGQLTVESITASPDLLKLLI 156
Db 107 TFMND-SYVVKSVYDSDSYNSVLEGGT-----FIHNGQLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQVHSPFADYVQDVESLKSVPMPQEGDIQYAAQSPVGNITAFVRENDLYW 155
Db 157 RTNSQVNRHSTFGSYFYVD-----KSSSFEEIGNEVALAIWSPNSNDIAYVDNNIYI 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPDWIYEEIILGDRYALWFGPDGEYLAFLSNFTGVP 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFPEDDKAAWWSPTGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPYWPWELKIRYKVSQTNPTVTLISLNIAASKEVKQAPIDAFEST-DL 271
Db 273 FIIPYVQDEK--DIYPEMRSIKYPKSGTPNPHAEALWVYSMDKGTGFPHRISGNKKGSL 330
Qy 272 IIGEVANLTDTHTTAAKAFNRVQDQKVAVDVASNKATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTTRDSSDLTLVFLDITIAKTSNVVRNESSNGGWETHNTLTF 388
Qy 332 IGPIKPSDKDAY-YIDISDHSQWHAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.  
  
OM protein - protein search, using sw model  
  
Run on: January 9, 2006, 03:21:01 ; Search time 19 Seconds  
(without alignments)  
337.510 Million cell updates/sec  
  
Title: US-10-723-807-2\_COPY\_17\_771  
Perfect score: 4033  
Sequence: 1 LDVPRKPHAPTEGSKRLTF.....EKNRQTQVLMHQWTKDLEE 755  
  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
  
Searched: 61141 seqs, 8493638 residues  
  
Total number of hits satisfying chosen parameters: 61141  
  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
  
Database : Published Applications AA New:\*  
1: /cgn2\_6/protdata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/protdata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/protdata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/protdata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
5: /cgn2\_6/protdata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
6: /cgn2\_6/protdata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
7: /cgn2\_6/protdata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
8: /cgn2\_6/protdata/1/pubpaa/US60\_NEW\_PUB.pep.\*

us-10-723-807-2\_copy\_17\_771.rapbn  
  
26 114 2.8 1250 7 US-11-137-465-62 Sequence 62, Appl  
27 111.5 2.8 1857 7 US-11-057-058-60 Sequence 60, Appl  
28 111.5 2.8 1857 7 US-11-057-058-61 Sequence 61, Appl  
29 111 2.8 1889 7 US-11-102-476-46 Sequence 46, Appl  
30 110.5 2.7 794 6 US-10-485-517-355 Sequence 355, App  
31 110 2.7 928 6 US-10-880-144-4 Sequence 4, Appl  
32 110 2.7 1225 7 US-11-102-476-2 Sequence 2, Appl  
33 110 2.7 2105 7 US-11-052-554A-173 Sequence 173, App  
34 109 2.7 715 7 US-11-052-554A-290 Sequence 290, App  
35 108.5 2.7 1254 6 US-10-528-031-47 Sequence 47, App  
36 108.5 2.7 1263 6 US-10-485-517-127 Sequence 127, App  
37 108 2.7 866 7 US-11-052-554A-298 Sequence 29, Appl  
38 106 2.6 1151 6 US-10-793-626-2448 Sequence 2448, Ap  
39 105 2.6 472 7 US-11-165-211-1 Sequence 1, Appl  
40 105 2.6 472 7 US-11-165-226-1 Sequence 8, Appl  
41 103.5 2.6 1881 7 US-11-040-472-8 Sequence 174, App  
42 103.5 2.6 2204 7 US-11-052-554A-174 Sequence 3, Appl  
43 103 2.6 472 7 US-11-165-211-3 Sequence 3, Appl  
44 103 2.6 472 7 US-11-165-226-3 Sequence 3, Appl  
45 102 2.5 472 7 US-11-165-211-2 Sequence 2, Appl

ALIGNMENTS  
  
RESULT 1  
US-10-522-789-2  
; Sequence 2, Application US/10522789  
; Publication No. US20050260732A1  
; GENERAL INFORMATION:  
; APPLICANT: TANABE SEIYAKU CO., LTD.  
; TITLE OF INVENTION: Three-dimensional structure of dipeptidyl peptidase IV  
; FILE REFERENCE: 03-039-PCT  
; CURRENT APPLICATION NUMBER: US/10/522,789  
; CURRENT FILING DATE: 2005-01-28  
; PRIOR APPLICATION NUMBER: US 60/398,761  
; PRIOR FILING DATE: 2002-07-29  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 766  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-522-789-2

Query Match 24.5%; Score 987.5; DB 6; Length 766;  
Best Local Similarity 32.6%; Pred. No. 3.3e-67;  
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;  
  
QY 11 TCGSKRLTFNFTVVKQAITPTSRVQWLSGAEDSGSVYAAEDGSLTIENIVTNESTLI 70  
DB 36 TADSRKTYLTLD-YLKNTRYLKLXLSRWIS---DHEYLKQENNILVFNAYCGNSVFL 91  
QY 71 PADKIPTGKEAFNYIHPDLSSVLWASNTKQVRSFFADYVQDVESLSKSVLMPDOE- 129  
DB 92 NSTDFBFGHSINDYSISPDGQFILLFVYVQWRHSYASY---DIYDLNKRQLITEERI 148  
QY 130 -GDIQYAOQSPYGNITAFVRENDLYV--WDNGVTVTRITDDGGPD-MFHGVPDVIYEEIL 185  
DB 149 PNTQVTVTPSPGHKLAYVWNNNDIYVKBPNLPSYRITWTGKEDIYNGITDMWYEEVF 208  
QY 186 GDRYALWFSPDGEBYLAYSFNETGVPTTYVQYMDNQEIAPAYPWLKIRYPKVSQNPRT 245  
DB 209 SAYSALWSPNGTFFLAYAQFNDEVEPLIEVFSYSDS---SLQPKTVRVVPYKAGAVNPT 265  
QY 246 VTLSLNI--ASKEVQKQPIDAFESTDLIG-----EVAWLTOHTTVAAKAFNRVQDQ- 297  
DB 266 VKFFVNTDLSLSSVNTATSIQTAPASMLIGHYLCDDVTWATQER--ISLQWLRRIQNY 323  
QY 298 -QKVAVDTASNKATVISDRD-----GTDGWLNDLLSMKXIGPKPS-----DKDAYID 346  
DB 324 VMDICDYDESSGRWNCILVARQHIEMSTTGM-----VGRFRPSEPHFTLDGNSFYKI 374

SUMMARIES  
  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
  
Result No. Score Match Length DB ID Description  
1 987.5 24.5 766 6 US-10-522-789-2 Sequence 2, Appl  
2 985 24.4 762 7 US-11-116-939-13 Sequence 13, Appl  
3 977 24.2 760 7 US-11-186-284-55 Sequence 55, Appl  
4 243 6.0 657 7 US-11-179-977-1 Sequence 1, Appl  
5 133.5 3.3 1424 6 US-11-102-476-4 Sequence 4, Appl  
6 131 3.2 668 6 US-10-454-437-118 Sequence 118, App  
7 123.5 3.1 621 6 US-10-467-657-1554 Sequence 1554, App  
8 123.5 3.1 1045 7 US-11-113-424-54 Sequence 54, Appl  
9 123.5 3.1 1094 6 US-10-821-234-1097 Sequence 1097, Ap  
10 123.5 3.1 2376 7 US-11-096-051-2 Sequence 4, Appl  
11 123.5 3.1 2715 7 US-11-096-051-4 Sequence 2, Appl  
12 123.5 3.1 2721 7 US-11-096-051-10 Sequence 10, Appl  
13 123.5 3.1 2725 7 US-11-096-051-8 Sequence 8, Appl  
14 122.5 3.0 1259 6 US-10-995-561-625 Sequence 625, App  
15 122.5 3.0 1286 6 US-10-995-561-628 Sequence 628, App  
16 122.5 3.0 1341 6 US-10-995-561-621 Sequence 621, App  
17 122.5 3.0 1348 6 US-10-995-561-624 Sequence 624, App  
18 122.5 3.0 2355 6 US-10-995-561-623 Sequence 623, App  
19 122.5 3.0 2355 6 US-10-995-561-627 Sequence 627, App  
20 122.5 3.0 2384 6 US-10-821-234-1545 Sequence 1545, Ap  
21 122.5 3.0 2386 6 US-10-995-561-626 Sequence 626, App  
22 120.5 3.0 2715 7 US-11-113-424-51 Sequence 51, Appl  
23 118.5 2.9 544 7 US-11-052-554A-358 Sequence 358, App  
24 117 2.9 1315 6 US-10-995-561-630 Sequence 630, App  
25 117 2.9 2296 6 US-10-995-561-633 Sequence 633, App

Qy 347 ISDSGWAHLVLPVSGGEPILTKGDWEVTSILSIDOEROLVYLLSTQHSTE--RHLY 404  
Db 375 ISNEGYRHICYFQIDKKDCTFTITGTWEVIGIEALTSD--YLYIISNEYKMGPGGRNLY 432  
Qy 405 SVSVSTFA-VTPL--VDDTVAAYWSASFSANSYVILTYGGPDVPYQELVYTTNSTKPLR 460  
Db 433 KIQLSDYTKVTCLSCELNPERCQYVSFSKEAKYQLRCSGPGPLPLYTLHSSVNDKGLR 492  
Qy 461 TITDNA---KYLEQIK-----DYALPNITYFELPSPGETLNMVQRLPPGFSPPDKKYP 511  
Db 493 VLEDNSALDKLVNQVQPSKKLDLFIILNETKFWQM-----ILPPHFDKSKKYP 542  
Qy 512 LFTYGGGAEVTKRQWALNPKAYVADSLELYVTWVNRGTGFKGRKRSVATVQLG 571  
Db 543 LLDVYAGCSQKADTVFR-LNWTATYLASTENI--IVASFDGRGSGYQGDKIMHAINRRLG 599  
Qy 572 LLEAEDQIYAAQAANIPWIDADHIGTWGSGFGLVTSKLEKDSGAFGLGVITAPVSDW 631  
Db 600 TFEVEDQLEAARQFSKMGFVNDKRIALMGWSYGGVTSWVLGSGGVFKCGIAVAPVSRW 659  
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIQHGTGDDNVHFQNSA 688  
Db 660 EYDVSVYTERYMGPTPEDNLDRNSTRNVTMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717  
Qy 689 ALVDLLMGDGVSPKLSHONFTSDHGISYHGGGVFLYKQLARKLYQ 735  
Db 718 QISKALVDVGVD---FQAMWYTDHGIASHTAHQIYTHMSHFQ 761

## RESULT 2

US-11-116-939-13  
; Sequence 13, Application US/11116939  
; Publication No. US20050265995A1  
; GENERAL INFORMATION:  
; APPLICANT: Stephen Tomlinson  
; APPLICANT: Richard J. Quigg  
; TITLE OF INVENTION: TISSUE TARGETED COMPLEMENT MODULATORS  
; FILE REFERENCE: 19113.0115U2  
; CURRENT APPLICATION NUMBER: US/11/116,939  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: 60/565,907  
; PRIOR FILING DATE: 2004-04-28  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 762  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; note=synthetic  
US-11-116-939-13

Query Match 24.4%; Score 985; DB 7; Length 762;  
Best Local Similarity 32.9%; Fred. No. 5e-67;  
Matches 245; Conservative 129; Mismatches 288; Indels 82; Gaps 26;

Qy 34 RSVWLGAEDGSVYAAADGSLTIENVTNESRTLIPADKIPTKEAFNWIHPDLSSV 93  
Db 54 KKLRLWIS---DHEVLYKQENLILVNAEYSSVPLENSTDFEGHSINDYSISPDGQFI 110  
Qy 94 LWASNHTKQYRHSFPADYVQDVESLKSVPMLPDQE--GDIQYQAWSVPGNTIAPVRND 151  
Db 111 LLEYNVYQWRHSYTSY---DIYDLNKRQLITEERIPNNTQWVTWSPVGHKLAYVMND 167  
Qy 152 LYV--WDNGVTTRITDDGPD-MFQGVDPDWIYEEBILDRVALWFSDPGEVLAYSNET 208  
Db 168 IYVKIEPLPSYRITWTGKEDIYNGITDWYVEEVFSAYSALMWSNPGTFLAYAFQNDT 227  
Qy 209 GVPTVTVOYMDNQBIAPAYPWLKIRYPKVSQNTPTVTLISLNI--ASKEVKQAPIDAF 266  
Db 228 EVPLIEYSFYSD---SLQYPKTVRPVPKAGAVNPTVKFVWNTDSLSSVTNATSIQIT 284

Qy 267 ESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ--QKVVAVDTSANKATVISDRD-- 317  
Db 285 APASMLIGDHYLDCVDTWATQER--ISLQWLRRIQNYVMDCIDYDESSGRWNCVIAQOHI 342  
Qy 318 --GTDGDLNLLSMKYIGPIKPS-----DKDAYIIDISDHSWAHLVLPVSGGEPIL 369  
Db 343 EMSTTGW-----VGRFSPSEPHEFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFI 393  
Qy 370 TKGDWEVTSILSIDOEROLVYLLSTQHSTE--RHLYSVSYSTFA-VTPL--VDDTVA 423  
Db 394 TKGWEVIGIEALTSD--YLYIISNEYKMGPGGRNLYKIQLSDYTKVTCLSCELNPERCQ 451  
Qy 424 YWSASFSANSYVILTYGGPDVPYQELVYTTNSTKPLATITDNA---KYLEQIK-----D 474  
Db 452 YYSVSFSKEAKYQLRCSGPGPLPLYTLHSSVNDKGLRVLENSALDKLVNQVQPSKKLD 511  
Qy 475 YALPNITYFELPSPGETLNMVQRLPPGFSPPDKKYPILFTYGGGAEVTKRQWALNPK 534  
Db 512 FIILNETKFWQM-----ILPPHFDKSKKYPILLDYYAGCSQKADTVFR-LNWA 560  
Qy 535 AYVASDSELYVTWVNRGTGFKGRKRSVATVQLGLEAEDQIYAAQAANIPWIDAD 594  
Db 561 TYLASTENI--IVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMGFVDNK 618  
Qy 595 HIGTWGSGFGLVTSKLEKDSGAFGLGVITAPVSDMRFYDSMYTERYM--KTLSTNEEG 652  
Db 619 RIATWGSYGGVTSWVLGSGGVFKCGIAVAPVSRWEYDVSVYTERYMGPTPEDNLH 678  
Qy 653 YETSAV-RKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHONFTD 711  
Db 679 YRNTVMSRAENFKQVE--YLLIHGTADDNVHFQNSAISKALVDVGVD---FQAMWYTD 733  
Qy 712 SDHGISYHGGGVFLYKQLARKLYQ 735  
Db 734 EDHGIASHTAHQIYTHMSHFQ 757

## RESULT 3

US-11-186-284-55  
; Sequence 55, Application US/11186284  
; Publication No. US20050266493A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berger, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Thibodeau, Stephen N.  
; APPLICANT: Burgart, Lawrence J.  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF COLON CANCER  
; FILE REFERENCE: MPM01-029P2RNM  
; CURRENT APPLICATION NUMBER: US/11/186,284  
; CURRENT FILING DATE: 2005-07-21  
; PRIOR APPLICATION NUMBER: US/10/301,822  
; PRIOR FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 60/339,971  
; PRIOR FILING DATE: 2001-12-10  
; PRIOR APPLICATION NUMBER: US 60/361,978  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/381,988  
; PRIOR FILING DATE: 2002-05-20  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 55  
; LENGTH: 760  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-11-186-284-55

Query Match 24.2%; Score 977; DB 7; Length 760;

```

Best Local Similarity 32.3%; Pred. No. 2e-66;
Matches 242; Conservative 140; Mismatches 293; Indels 74; Gaps 25;

QY 4 PRKHAPTEGSGKRLT-----FNETVVKQALITPTSRVOWLSGAEDGSGVYAAEDGSLTIE 59
DB 28 PSRVHNSSENTMRALTLLKDLINGTFSYKTFPP-----NWSIQE---YLHOSADNNVLY 79
QY 60 NIVTNESRTLPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESL 119
DB 80 NIETQGSSTIL-SNRTMKSVAASNYGLSPDRQFVLESDYSKLYWYSTAYIYIDLNG 138
QY 120 KSVPLMPDQEGD-----IQYAWQSPVGNITAFVRENDLYWMD--NGTVTRITDDGPD-M 171
DB 139 EFV-----RGNELPRPIQYLCWSVPVSKLAIVYQNNIYKQRPDPFPQITFNGRENKI 192
QY 172 FHGVPDNIYEBEILDRVALMFPSPGGEYLAYSFNETGVPTVYQYMDNQEIAPYWE 231
DB 193 FNGIPDWYEEEMLPKTYALMWSNPKFLAYAEFNDKQIPVIAYSYYGDEQ-----YPR 247
QY 232 LKIRYPKVSQTNPTVTLN-----TASKEVKQAPIDA-PESTDLLIIEVAVLTDTH 284
DB 248 INIPPKAGAKNPPVRIIFIDTTPAYGQEV---PVPAMTASDDYFSLTWTDTER- 303
QY 285 TVAAKAFNRVQDQKVAVD-----TASNKATVISDRDGTGDLNLLSMKYIGPIKPS 338
DB 304 -VCLQWLXRVQNVSLICDFREDQWDCPKTQEHIEBSRTGWAGGFVSR---PVFSY 359
QY 339 DKQAYIIDSHSGWAHLYLPVPSGGEPIPLTKGDWEVTSILSIDQERQLVYVLT--QH 396
DB 360 DAISYKIFSDKDGKHIYIKDTVENAIIQTSKWEAINIPRVTDQS--LFYSSNEFEE 417
QY 397 HSTERHLYSVSSTF-----ATPLVDDTVAAWASFSANSYGYILTYGGDPVPQELY 451
DB 418 YPGRENIYRISGSGYPPSKKCVTKLXERCOYITASFSYAKYALVCYGGIPISTUH 477
QY 452 TTNSKPLRTIDNAKLEQIKDVALPNITYFELPLPSGETLNNMQLPPGSPDKYPI 511
DB 478 DORTQEIKILEENKELENALKNIQLPKEIKLEVE-ITLWYKMLPPQDRSKYPL 536
QY 512 LFTYGGGQAQVTRKQWALNFKAYVASDSELEYVTWVDNRGTGPKGRKFRSAVTRQLG 571
DB 537 LIQVGGPCQSQRVSF-AVNWISYLASKEGM--VIALVDRGRTAFQGDKLLYAVYRKIG 593
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGIWMGSGGYLTSKLEKDSGAFTLGVITAPVSDW 631
DB 594 YVEVEDQITAVRKFTEMGFIDKRIALWGSYGGIVSSIALASGTLFCGIAVAPVSW 653
QY 632 RFYDSMYTERYM--KTLSTNEEGYETSAY-RKTDGFKNVEGGFLIOHGTGDDNVHFNQSA 688
DB 654 EYASVYTERFGLPTKDDNLEHYKNSVTWAAEYFRVD--YLLIHGTADDNVHFNQSA 711
QY 689 ALVDLLMGDGVSPKLSQWFTSDSHGIS 717
DB 712 QIAKALVNAQVD---FOAMWYSQNHGLS 737

```

```

RESULT 4
US-11-179-977-1
; Sequence 1, Application US/11179977
; Publication No. US20050249789A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/11/179,977
; PRIOR FILING DATE: 2005-07-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-11-179-977-1

```

```

Query Match 6.0%; Score 243; DB 7; Length 657;
Best Local Similarity 20.9%; Pred. No. 6.6e-11;
Matches 146; Conservative 113; Mismatches 333; Indels 106; Gaps 26;

QY 67 RTLIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLKSVPMLP 126
DB 2 KKLITADDITAVSVTDPOYADGTFRAYVKSQVQNEKDSYTSNIWITETKTGGSVPTW 61
QY 127 DOEGDIQYAWQSPVGNITAFV--REND---LYVM--DNGTVTRITDDGPDPMFHGVPDW 179
DB 62 GEKSTQ-PRNSPDGRTLAFISDRGDAALYINSTEAGEARKLTD-----IP--- 108
QY 180 YEEIILDRYALMFPSPDGE-YLAYSFNETGVPTVYQYMDNQEIAPYWEIKIRYPK 238
DB 109 -----YGVSKPLW-SPDGEISILVTSIGEGESIDDRKTEQDSYEPVEVQGLSYK-RD 161
QY 239 VSOQNTPTVTLNLLNATASKEVKQAPIDAPESTDLLIIEVAVLTDTHTTVAAKAFNRVQ 298
DB 162 GLTRGAYAAQLVLVSVKSGEMKE-----LTSHKADHGDPAFSPDGKWLVPFSA 216
QY 299 K---VAVVDTASNKATVISDRDGTGDLNLLSMKYIGPIKPSDKD-----AYYIDI 347
DB 217 KPHDVIMLSGDLKQVTPHRSFGSSFSFDPDGYLA-LLGNEKEYKNATLSKAWLYDI 275
QY 348 SD-----HSGWAHLYLPVPSGGEPIPLTKGDWEVTSILSIDQERQLVYVLTQ 395
DB 276 EQGRILTCLTEMLDVHLADALIGDSLIGGAEQRPITWKDSQGFYVIGTDQSGTGIIYI 332
QY 396 HSTERHLYSVSSTFATVPLVDDTVAAWASFSANSYGYILTYGGDPVPQELYTTNS 455
DB 333 --SIEGLVYIRLEKEYIN-----SFLSPDEQHFITASVTKPDRP-SELYSIPL 378
QY 456 TKPLRTITD-NAKLEQIKDVALPNITYFELPLPSGETLNNMQLPPGSPDKYPIPLFT 514
DB 379 GQEEKQLTGANDKF---VREHTISIPBEIYQATGDEGVWVNGWLMRPAQNEGETTYFLILN 435
QY 515 PYVGGP---AQEVTKRQWALNFKAYVASDSELEYVTWVDNRGTGPKGRKFRSAVTRQLG 571
DB 436 IHGGPHMVMYGHYTFYHEFQVLAAGYAVV-----YINPRGSHGYGQEFVNAVRGDY 486
QY 572 LLEAEDQIYAAQAA-ANIPWIDADHIGIWMGSGGYLTSKLEKDSGAFTLGVITAPVSD 630
DB 487 GKDYDDVMQAVDEAIKRDPIHIDPKRLGVGTGSGYGFMTNWIQV--TNRFKAAVTRQSI 545
QY 631 RFYDSMYTERYMKT-----LSTNEEGYETSAYRKTGDKNVEGGFLIOHGTGDDNV 682
DB 546 WISFHVGSIDIGYFTTDWQLEHDMFDETEKLWDRSPKLYA---ANVETPLILLHGERDRC 602
QY 683 HFQNSAALVDLLMGDGVSPKLSQWFTSDSHGISYHG 720
DB 603 PIEQAEQLFIALKKWQKTKLVR---FPNASHNLSRTG 637

```

```

RESULT 5
US-11-102-476-4
; Sequence 4, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U.M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102,476
; PRIOR FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4

```

```
; LENGTH: 1424
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-4

Query Match      3.3%; Score 133.5; DB 7; Length 1424;
Best Local Similarity 19.0%; Pred. No. 0.04;
Matches 164; Conservative 103; Mismatches 307; Indels 291; Gaps 38;

Qy 28 AITPSSVQWLS-----GAEDGSVVAADGSLTIENIVNESRTL-----IPADKI 75
Db 92 SIVPDSQSVTTIQNRVGRGIAGSSIIKAEVNGLYSEQKIVTPTATLNSIQVTSLESGIL 151
Qy 76 PTGKE-----AFNYIWIHPLDLS---VLWASNH-----TKQYRHSF 107
Db 152 PKGTNRQFSAIGFSDGSHQDISNDPLIVWSSSNPDLVQVDDSLASGLNLTGTAHIRASF 211
Qy 108 -----PADYVQDVESLKSVP LMP-----DQGDIOYAO--WSPVG 141
Db 212 QSKQGAEMTVGDAVLSQIQVTSNPNIPGLKKQKLIATGIYSDNSNRDISSSVIMNSN 271
Qy 142 NTIAFVREND--LYVMDNGTVTRITDDGDPDMFHGVPDVIYEEELGD-----RYAL 191
Db 272 STIANQNGILETADGIVTISAS-----SENIGSKLIVTPAALVSI 316
Qy 192 WFPSDGEYLA---YLSFNETGVPTTYVQYVMDNQEIAPAYPWE-----LKIRYPK--- 238
Db 317 SVSPNTSVAKGLQENFKATGIFTN-----SNSDITDQVTDWSSNTDILSISNASDSHG 371
Qy 239 -----VSQTNPTVTLSLNIASK---EVKQAPIDAFESTDLIIGEVAMITDTHTTVAAKAF 291
Db 372 LASTLNQGNKVTASIGGIQGSTDTFTVQAALTSEIVSPVLPSIAKGLTQKFTAIGIFTD 431
Qy 292 NRVDQKQV-----AVDTASN-----KATVISDRCT-----DGMV-----DNL 326
Db 432 NSKKDITDQVTDWSSSALVSVNSNDNKKGLKAHAGVDTTITATLGVKSGKTLVTVPAV 491
Qy 327 LSMKYIGIPKSPDKDAYVIDISDHSGWAHLVLPFVSGGEPIPLTKG---DWEVTSILSID 383
Db 492 LTSIQINPVNS-----LAKGLTKFSAIGIYSDN 521
Qy 384 QEROLVYLLSTQHHSSTERHLYSVSYSTPAVTPVDDTVAAVWSASFANSYIILTVGGP 443
Db 522 SNKDI-----TSAVTFSSSDSIATISNAQKNQGNAYGAATGAT 560
Qy 444 DVPQELYTNSTKPLRTITDNAKVLQI KDVALPNITYFELPLPSGETLNVQMRLPGF 503
Db 561 DI---KATFGKVSPSVTSLSTAACLVEIQ----- 587
Qy 504 SPDKYPILFTPYGGPGAQEVTKRWQALNFKAYVASDSELEYVTWTVDNRG-----TGPKG 559
Db 588 -----ITPAAKAKAGLTERFKATGIFTDSSNDITNQVTWSSNTDIAEITNSG 638
Qy 560 RKFRSAVTQGLLEAEQOIVAAQQAANI PMIDADHIGWSPGGYLTSKVLEKDSGAF 619
Db 639 SK--GITNT--LTPGSEISAA-----LGSIKSSKVLKVTAPAQ 673
Qy 620 TLGVITAPVSD-----WRFYDS--MYTERYMKTLSTNEEGVETSAVRKTDGFKNVEGGF 671
Db 674 LISIAVTPIINSAVKGLIRQKATGTYTDHSVQDV--TALATWSSSNPRK--AMNVNVTGS 730
Qy 672 LIQHTGDDNVHFQNSAALVDLLMGD---GVSPEKLHSQWFTSDPHGISYHGGGVFLYKQ 728
Db 731 VTTVATGNTNI-----KATIDSIGSSVLNVTPALLTSEIPTINSIT-HG-----LTQK 780
Qy 729 L-ARKLYQEKNRQTVLHMWTKKD 752
Db 781 FKATGIFSDKSTQNLTLQVLTWISSD 805

RESULT 6
US-10-454-437-118
; Sequence 118, Application US/10454437

Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 118
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-118

Query Match      3.2%; Score 131; DB 6; Length 668;
Best Local Similarity 17.8%; Pred. No. 0.021;
Matches 136; Conservative 100; Mismatches 270; Indels 268; Gaps 34;

Qy 78 GKEAFNYWIHPDLSSVLWASNHKQYRHSFPADYVYVQDVESLKSVP LMPDQEGDIQYAO 137
Db 59 GEKLYNFRDAQHPRGVWRITTLSEYE---SDQPEWDV--LIDVDALAEDEG----- 105
Qy 138 SPVGNITAFVRENDLYVMDNGTVTRITDDGDPDMFHGVPDVIYEEELGDRYALWFSPDG 197
Db 106 -----EN--WVM-KGAVVR-----SPEF 120
Qy 198 EYLAVLSFNETGVPTTYVQYVMDNQEIAPAYPWEKIRYKPVSTQNTPTVTLISLNIASKE 257
Db 121 D-RALVFRSGGADATVIREF-----DLATAAFVDDSPFE 154
Qy 258 VKQAPIDAFESTDLIIGEVAMITDTHTTVAAKAFNRVQDQKVVAVDTASNAKATVISDRD 317
Db 155 LKEAKSD-----VTWV-DLDTL-----LVGTDGEGSLT----- 182
Qy 318 GTDGLWNLNLSMKYIGIPKPSDK--DAYVIDISDHSGWAHLVLPFVSGGEPIPLTKGDWE 375
Db 183 -DSGYPARVLTVKRGTPLEQAELFPEGSRQDVATHA-WR-----DSTPG-FE 226
Qy 376 VTSILSIDQERQLVYLLSTQHHSSTERHLYSVSYST-----FAVTPLVDDTVAAAYW 425
Db 227 RTFV-----SRSLDFYNSSETLETGGLVLDVPTDCDVIVKQWIFVSPRTD-----F 275
Qy 426 SASFSANSQYVIL---TYGGDPVYPQELTYTNSTKPLRTI---TONAKVLEIQIKYAL--- 477
Db 276 AGIPAGGLGVLLKFELEGGRD--FQPVFTPTTESTSLQGLATTKNFLVTLNNVSTRIV 333
Qy 478 -----PNITYFELPLPSGETLNVW-----QRLPPGFSPPD 506
```



Db 334 TVPLNDPTEHEHIDLPEHVTAHVATGSLDGEIWOQAASFTEAPTILLRAELPGALBAV 393  
Qy 507 KYPILFTPGGPGAGQEVTKW-----QALNFKAY-----VA 538  
Db 394 KKAQLOFENAG-----QETROHWATSADGKIPYFITGAFEBEPQNTLVHAYGGFEVSLTP 449  
Qy 539 SDSELEYVTW-----TVDNRGTFKGRKFRSAVTRQLGLEAEADQIYAAQQAANIPW 590  
Db 450 SHSPTRGTAWLEKGYFYVEANLRGGEGEPWHSSQATKLNRMKWEDHRAVLADLVERGY 509  
Qy 591 IDADHIGWMSFGYLSKVLKSDGAPFTGLVITAPVSDM-RFYDSMYTERYMKTLSTN 649  
Db 510 ATPEQIAIRGGSGGLLTSGALTQYPAFGAAVQVPLADMLRYHTWSAGASWMAEYGNP 569  
Qy 650 BEGYETSAVRKTDFKNVEG-----GFLIOHGTGDDNVHFNQNSAALVDLLMGDGVSP 702  
Db 570 DDPEERAVIEQYSPVOAVGVGVEKRYPPALVTTSTRDRVHPAHARLFAQALLDGAQAVD 629  
Qy 703 KLHSQWFTSDSHGISHYGGVFLYKQLARKLYQEKNRQT---QVLMHQWTKKDLLE 755  
Db 630 -----YENTEGG---HAGAA-----DNKQTAFFVESLIYTWIEKTLQ 664

## RESULT 7

US-10-467-657-1554  
; Sequence 1554, Application US/10467657  
; Publication NO. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON Spa  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 1554  
; LENGTH: 621  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-1554

Query Match 3.1%; Score 123.5; DB 6; Length 621;  
Best Local Similarity 20.1%; Pred. No. 0.069;  
Matches 134; Conservative 83; Mismatches 226; Indels 225; Gaps 37;  
Qy 129 EGDIOAOWSPVG-NTIAFVRENDLYV---WNGTWTTRITDGGPDMFHGVPDMEYEEI 184  
Db 104 EGGFHF---PAGKHNSWRDENSVMVCPAWE---RQLTESG-----YPREV 144  
Qy 185 LGDRVALWSPDGEYLAYSFNETGYPTTVQ-----YMDNQETAPAYPWELKIR 235  
Db 145 -----WLVERGK-----SPES-LPAYOIDKGMVMNARYLDPQ----- 178  
Qy 236 YPKVSTQNTPTVTLSSLNIASKEVKQAPIDAFESTDLIIGEVAWLTDTHTTVA---AKAF 291  
Db 179 -----GSPIDLEASD-----GFYTKYLVQSSEGGAKPL 208  
Qy 292 NRVDQOKVAVDVTASNKATVISDRDGTGWLDN---LLSMKYIGPIKPSDKXAYIDISD 349  
Db 209 NLPNDQDV-----GYLAGHLLTLR-----KDWHRANQS- 238  
Qy 350 HSGWAHLYLFPVGGGEPILTKGDWEVTSILSDQERQLVYVLSQHHSTERHLYSVYS 409  
Db 239 -----YPSGALVAVKLNRGELGAAQLLPAFDETOALESVET-----TKR----- 277  
Qy 410 TFAVTPLVDDTVAAVWSAFSFSAN-----SGYIILT---YGGPDVPVQELVTTNS 455

Db 278 -FVVASLLENVQGRKAWRFADSKWQBAELPHLPSGALEMTDQPMGG-DVVY--LAASDF 333  
Qy 456 TKPLRTITDNKVLK-----QIKDYALPNITYFELPLPSGETLLVMQRLP-----PGFS 504  
Db 334 TTPLTLFALDNLNVELTVMRLQPOQFVSDGIEVRQFMAVSSDG-----ERIPYFHVGNAA 389  
Qy 505 PDKKYPILFTPYGSGAQE-----VTKRQOALNFKAYVASDSSELEYVTWTVDRNGTGF 557  
Db 390 PDT--PTLVAYAGGFGIPELPHYLGSVGKYWLE-EGNAFVLANI-----RGGE 435  
Qy 558 KGRKFRSAVTRQLGLEAEADQIYAAQQAANIPWIDADHIGWMSFGGYLSKVLKSDG 617  
Db 436 FGPRHQAAQGISKHKSVDDLLAVRDJSEGMSSPKHIGLOGGNGGLUITAAAFVREPO 495  
Qy 618 AFTLQVITAPVSDWRFY-----DSMYTBRYMKTLSTNEEGYETSAVR--KTDFGKNVEGG 670  
Db 496 SIGALVCEVPLTDMIRYPLLSAGSSWTDEY-----GNPKYEACKRRRLGELSYPHNLSDG 550  
Qy 671 F-----LIHGTGDDNVHFNQNSAALVDLLMGDGVSPKLSQWFTSDSHGISHYGGVFL 725  
Db 551 IDYPPALLITSLSDRVRHPAHALKFYAKLR--ETSPQ-----SWLYSPDGG-GHTGNGT-- 601  
Qy 726 YKQLARKL 733  
Db 602 QRESADKL 609

## RESULT 8

US-11-113-424-54  
; Sequence 54, Application US/11113424  
; Publication NO. US20050260713A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli et al.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-225  
; CURRENT APPLICATION NUMBER: US/11/113,424  
; CURRENT FILING DATE: 2005-04-21  
; PRIOR APPLICATION NUMBER: 60/256,704  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/311,590  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/257,314  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/311,613  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/315,617  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/307,506  
; PRIOR FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: 60/322,358  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/294,075  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/288,153  
; PRIOR FILING DATE: 2001-05-02  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54  
; LENGTH: 1045  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-113-424-54

Query Match 3.1%; Score 123.5; DB 7; Length 1045;  
Best Local Similarity 20.0%; Pred. No. 0.15;  
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;  
Qy 60 NIWT--NESRTIIPADKIPTGKEA-ENYWIHPDLSSVLWASNHTKQYRHSFFADYVQDV 116  
Db 269 SIITDNEGLLLQTAFLGTSTRVLFKYRRQRTLSILYDST-----RVSF---TYDETA 320  
Qy 117 ESKSVPLMPDQEG---DIQYQWSPVGNITAF-----VRENDLYVWDNGTGTTRITDD 166

```
Db 321 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIRFSESDGMVNAFSDYDNS--FRVTSM 376
Qy 167 GGPDMFHGVPDWIYEEIIGDRYALWFSPOGEY-LAYLSFNE---TCVPTYT-----214
Db 377 QGVINETPLPIDLYQFDDISGKVEQF---GKFGVIYDINQIISTAVMTYTKHFDAGR 432
Qy 215 ---VOYMDNQEIAPAYPWEIKIRYPKVSOTNPVTLSLNIASKEVQKAP-----262
Db 433 IKEIQY-----EIFRSLMYWITIYDNMGR-----VTKREIKIGFPANTTKVAY 476
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISD---R 316
Db 477 EYDVGQQLQTVYLNKIMRWNYDLNGLNHLNPSNARLTPLRYDLRDRITRLGDVQYR 536
Qy 317 DGTGWL-----DNLISMKYIGIPKPSDKDAYIIDISDHSGMAHLXLPVPSGG 364
Db 537 LDEGFLRQGRTEIFEYSSKGLLTVY-----SKSGWTVIYRYDGLGR 580
Qy 365 EPIPLTKGDMVEVTSILSDQERQLVYVYLSLTOHHSSTERHLYSVSYSTFAVTPLVDDTVAAY 424
Db 581 -----RVSKTSLGQHLQF-FYADLTYPTRITHVY--NHSSEITSLYYDLQHL 627
Qy 425 WSASFANSGYIITYLGGPDVPYQELYTNTSTKPLRTITDNKAKVLEIQIKOYALPNITYFE 484
Db 628 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQYTAYGEI-YFD 673
Qy 485 LPLPSGETLNVQRLPPGFS-----PDKKYPILFTPYGGPGAEVTKR--WQ 529
Db 674 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERYDILAGRWTTTPI--EIWKRIKGD 724
Qy 530 ALNFKAYVADSE-----LEYVT---WTV-----DNRGTGFKRKFERSAVTRQLG 571
Db 725 PAPNLNMFNRNNPASKIHDVKDIITDVSMLVTFGPHLHNAIPGFPVPKF-----D 776
Qy 572 LLEAEDQIYAAQQAANIPWIDADHIGI-----PDKKYPILFTPYGGPGAEVTKR--W 599
Db 777 LTPESYELVKSQQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGQSW 832
Qy 600 GWSFGGYLTSK-----VLEKDSGAFTLGVI-----TAPVSDWRFY-----D 635
Db 833 LW-----FATVKSLLTGKGVMLAVSQGRVQTNVLNANEDCIKVAALVNNAFYLENLHFTIE 888
Qy 636 SMYTERYMKTLSTNEEGYETSAAVKTGDKFNVEGGFLIQHGTGDDNHFQNSAALV-----691
Db 889 GKOTHYFIKT--TTPES-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 936
Qy 692 ----DLIMGDGVSPKLSHQWFTSDSHGISYHGGVFLYKOLARKLYOEKKNRQTOVLMHQ 747
Db 937 RRFADVENQFGAL--ALHVRV-----GMTLDEEKARILEQARQ---ALARA 978
Qy 748 WTK-----KDL EE 755
Db 979 WAREQQRVRDGE 991
```

## RESULT 9

```
US-10-821-234-1097
; Sequence 1097, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
```

```
; SEQ ID NO 1097
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1097

Query Match      3.1%; Score 123.5; DB 6; Length 1094;
Best Local Similarity 20.0%; Pred. No. 0.16;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

Qy 60 NIVT--NESRTLLPADKIPGKEA-FNYWIHPDLSVWNASNHTKQYRHSHFFADYYQDV 116
Db 318 SIITDYNEGLLQTAFLGTSRRVLFKYRQTRLSLSELYDST-----RVSF---TYDETA 369
Qy 117 ESLKSVPLMPDQEG---DIQYAQWSPVGNITAF-----VRENDLYWMDNGTVTRITDD 166
Db 370 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIRFSESDGMVNAFSDYDNS--FRVTSM 425
Qy 167 GGPDMFHGVPDWIYEEIIGDRYALWFSPOGEY-LAYLSFNE---TCVPTYT-----214
Db 426 QGVINETPLPIDLYQFDDISGKVEQF---GKFGVIYDINQIISTAVMTYTKHFDAGR 481
Qy 215 ---VOYMDNQEIAPAYPWEIKIRYPKVSOTNPVTLSLNIASKEVQKAP-----262
Db 482 IKEIQY-----EIFRSLMYWITIYDNMGR-----VTKREIKIGFPANTTKVAY 525
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISD---R 316
Db 526 EYDVGQQLQTVYLNKIMRWNYDLNGLNHLNPSNARLTPLRYDLRDRITRLGDVQYR 585
Qy 317 DGTGWL-----DNLISMKYIGIPKPSDKDAYIIDISDHSGMAHLXLPVPSGG 364
Db 586 LDEGFLRQGRTEIFEYSSKGLLTVY-----SKSGWTVIYRYDGLGR 629
Qy 365 EPIPLTKGDMVEVTSILSDQERQLVYVYLSLTOHHSSTERHLYSVSYSTFAVTPLVDDTVAAY 424
Db 630 -----RVSKTSLGQHLQF-FYADLTYPTRITHVY--NHSSEITSLYYDLQHL 676
Qy 425 WSASFANSGYIITYLGGPDVPYQELYTNTSTKPLRTITDNKAKVLEIQIKOYALPNITYFE 484
Db 677 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQYTAYGEI-YFD 722
Qy 485 LPLPSGETLNVQRLPPGFS-----PDKKYPILFTPYGGPGAEVTKR--WQ 529
Db 723 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERYDILAGRWTTTPI--EIWKRIKGD 773
Qy 530 ALNFKAYVADSE-----LEYVT---WTV-----DNRGTGFKRKFERSAVTRQLG 571
Db 774 PAPNLNMFNRNNPASKIHDVKDIITDVSMLVTFGPHLHNAIPGFPVPKF-----D 825
Qy 572 LLEAEDQIYAAQQAANIPWIDADHIGI-----PDKKYPILFTPYGGPGAEVTKR--W 599
Db 826 LTPESYELVKSQQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGQSW 881
Qy 600 GWSFGGYLTSK-----VLEKDSGAFTLGVI-----TAPVSDWRFY-----D 635
Db 882 LW-----FATVKSLLTGKGVMLAVSQGRVQTNVLNANEDCIKVAALVNNAFYLENLHFTIE 937
Qy 636 SMYTERYMKTLSTNEEGYETSAAVKTGDKFNVEGGFLIQHGTGDDNHFQNSAALV-----691
Db 938 GKOTHYFIKT--TTPES-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 985
Qy 692 ----DLIMGDGVSPKLSHQWFTSDSHGISYHGGVFLYKOLARKLYOEKKNRQTOVLMHQ 747
Db 986 RRFADVENQFGAL--ALHVRV-----GMTLDEEKARILEQARQ---ALARA 1027
Qy 748 WTK-----KDL EE 755
Db 1028 WAREQQRVRDGE 1040

RESULT 10
US-11-096-051-4
```

Sequence 4, Application US/11096051  
Publication No. US20050244868A1  
GENERAL INFORMATION:  
APPLICANT: Kekuda, Ramesh  
APPLICANT: MacLachlan, Timothy K  
APPLICANT: Rastelli, Luca  
APPLICANT: Vernet, Corine  
APPLICANT: Ettenberg, Seth  
TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
FILE REFERENCE: Attorney Docket No. Cura 967  
CURRENT APPLICATION NUMBER: US/11/096,051  
CURRENT FILING DATE: 2005-03-30  
PRIOR APPLICATION NUMBER: 10/038,854  
PRIOR FILING DATE: 2001-12-31  
PRIOR APPLICATION NUMBER: 10/455,772  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: 60/557,978  
PRIOR FILING DATE: 2004-03-30  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: CuraSeqList version 0.1  
SEQ ID NO 4  
LENGTH: 2376  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-096-051-4

Query Match 3.1%; Score 123.5; DB 7; Length 2376;  
Best Local Similarity 20.0%; Pred. No. 0.47;  
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY 60 NIWT--NESRTLPADKIPTGKEA-FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDV 116  
DB 1600 SIITDYNESGLLQTAFLCTSRRLVKYRRQTFELSEILDST-----RVSF---TYDETA 1651  
QY 117 ESKSVPLMPDQEG---DIQYQMSVGNNTIAP-----VRENDLYVWDNGTTRITDD 166  
DB 1652 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIFRSEDGMVNAFSDYDNS--FRVTSM 1707  
QY 167 GGPDMFPHGVDPDIYEEILGDRVALWFSPDGEY-LAYLSFNE---TGVPPTYT----- 214  
DB 1708 QGVINETPLFDLYQDDISGKVEQF-----GKFGVIYYDINOIISTAVMTYTKHFDAGH 1763  
QY 215 ---VOYMDNQEIAPYPMELKIRYPKVQSTNPTVTLSSLNIAASKVEKQAP----- 262  
DB 1764 IKEIQY-----EIFRSLMYWITIYDYMGR-----VTKREIKIGFPANTTKYAY 1807  
QY 263 ----IDAFESTDLIIGEVAMLTDTHTTAAKAFNRVQDQKVAADVTASNKATVSD---R 316  
DB 1808 EYDVEDQLOQTVYLNKIMWRNYDNLGNLHLLNPSNSARLTPLRYDLDRITRLGQVYR 1867  
QY 317 DGTGHWL-----DNLLSMKYIGIPKPSDKDAYVIDISDHSGWAHLVLPFVSGG 364  
DB 1868 LQBDGFLRQRGTEIFBYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 1911  
QY 365 EPIPLTKGDWEVTSILSDQERQLVYLSLTHHSTHERHLYSVSYTFAVTPLVDDTVAAY 424  
DB 1912 -----RVSSKSLGHLQF-PYADLTYPTRITHYV--NHSSEITSLYDYLQCHL 1958  
QY 425 WSASFSGANGYITLYTGGPDVPVQELTYTNSTKPLRTITDNKVLQIKDYALPNTIYPE 484  
DB 1959 FAMEISSGDEFYIA-----SDNTGTPFLAVFSSNGLMLKQIQVATYGEI-YFD 2004  
QY 485 LPLPSGETLNMORLPPGFS-----PDKKYPIILFTPYGGPGAEVTKR--WQ 529  
DB 2005 -----SNIDFQLVIGFHGGYDPLTKLIHFGERDYDILAGRWTTTPII-EIWKRIKGD 2055  
QY 530 ALNFKAVASDSB-----LEVVT-----WTV-----DNRGTGFKGRKFRSAVTRQLG 571  
DB 2056 PAPFNLYMFRNNPASKIHVDKDYITDVSNLVTFGPHLHNAIPGPPVKE-----D 2107  
QY 572 LLEAEQDIYAAQQAANIPIWDADHIGI-----FGVQOQVARQAKAFLSLGKVAEQVSSRRRAGGAQSW 599  
DB 2108 LTPESYELVKSQQWDDIPPI-----FGVQOQVARQAKAFLSLGKVAEQVSSRRRAGGAQSW 2163

QY 600 GWSFGGYLTSK-----VLEKDSGAFTLGV-----TAPVSDMRFY-----D 635  
DB 2164 LW-----FATVKSILGKGVMLAVSQGRVQTNVLNINIANEDCIKVAALVNNAFYLENLHFTIE 2219  
QY 636 SMYTERYMKTLSTNEEGYETSVAVRTDGFKNVEGGFLIQHGTGDDNVHVFQNSAALV---- 691  
DB 2220 GKOTHYFIKT--TTPES-DLGTFLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 2267  
QY 692 ----DLLMGDGVSPKHLHSQWFTSDHGISVHGGGVFLYKQLARKLYQEKNRQTOVLMHQ 747  
DB 2268 RRFADVEMQFGAL--ALHVRV-----GWLDEEKARILEQARQR---ALARA 2309  
QY 748 WTK-----KDLSE 755  
DB 2310 WAREQQRVRDGE 2322

RESULT 11  
US-11-096-051-2  
Sequence 2, Application US/11096051  
Publication No. US20050244868A1  
GENERAL INFORMATION:  
APPLICANT: Kekuda, Ramesh  
APPLICANT: MacLachlan, Timothy K  
APPLICANT: Rastelli, Luca  
APPLICANT: Vernet, Corine  
APPLICANT: Ettenberg, Seth  
TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
FILE REFERENCE: Attorney Docket No. Cura 967  
CURRENT APPLICATION NUMBER: US/11/096,051  
CURRENT FILING DATE: 2005-03-30  
PRIOR APPLICATION NUMBER: 10/038,854  
PRIOR FILING DATE: 2001-12-31  
PRIOR APPLICATION NUMBER: 10/455,772  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: 60/557,978  
PRIOR FILING DATE: 2004-03-30  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: CuraSeqList version 0.1  
SEQ ID NO 2  
LENGTH: 2715  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-096-051-2

Query Match 3.1%; Score 123.5; DB 7; Length 2715;  
Best Local Similarity 20.0%; Pred. No. 0.57;  
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY 60 NIWT--NESRTLPADKIPTGKEA-FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDV 116  
DB 1939 SIITDYNESGLLQTAFLCTSRRLVKYRRQTFELSEILDST-----RVSF---TYDETA 1990  
QY 117 ESKSVPLMPDQEG---DIQYQMSVGNNTIAP-----VRENDLYVWDNGTTRITDD 166  
DB 1991 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIFRSEDGMVNAFSDYDNS--FRVTSM 2046  
QY 167 GGPDMFPHGVDPDIYEEILGDRVALWFSPDGEY-LAYLSFNE---TGVPPTYT----- 214  
DB 2047 QGVINETPLFDLYQDDISGKVEQF-----GKFGVIYYDINOIISTAVMTYTKHFDAGH 2102  
QY 215 ---VOYMDNQEIAPYPMELKIRYPKVQSTNPTVTLSSLNIAASKVEKQAP----- 262  
DB 2103 IKEIQY-----EIFRSLMYWITIYDYMGR-----VTKREIKIGFPANTTKYAY 2146  
QY 263 ----IDAFESTDLIIGEVAMLTDTHTTAAKAFNRVQDQKVAADVTASNKATVSD---R 316  
DB 2147 EYDVEDQLOQTVYLNKIMWRNYDNLGNLHLLNPSNSARLTPLRYDLDRITRLGQVYR 2206  
QY 317 DGTGHWL-----DNLLSMKYIGIPKPSDKDAYVIDISDHSGWAHLVLPFVSGG 364  
DB 2207 LQBDGFLRQRGTEIFBYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 2250

Qy 365 EPIPLTKGDWEVTSILSIDQERQLVYVYLSLSTQHHSRHLHSVSVSTFAVTPLVDDTVAAV 424  
Db 2251 -----RVSSKTSILGQHLQF-FYADLTYPTRITHVY--NHSSEITSLSYLDLQHL 2297  
Qy 425 WSASFANSYGYILTYGGPDVPYQELVYTNSTKPLRTITDNKVLKQIKOYALPNITYFE 484  
Db 2298 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQTAYGEI-YFD 2343  
Qy 485 LPLPSGETLNMORLPPGFS-----PDKKYPILFTPYGPGGAQEVTKR--WQ 529  
Db 2344 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTPDI-EIWKRIKGD 2394  
Qy 530 ALNFKAYVASDSE-----LEYVT---WTV-----DNRGTGFKGRKFRSAVTRQLG 571  
Db 2395 PAPENLYMFRNPNASKIHDVKDYITDVNSLWTFGFHLHNAIPGFPVPKF-----D 2446  
Qy 572 LLEAQDIYAAQAANIPWIDADHIGI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGGAQSW 599  
Db 2447 LTPESYELVKVSSQWDDIPPI-----SDNTGTPLAVFSSNGLMLKQIQTAYGEI-YFD 2502  
Qy 600 GWSFGGYLTSK-----VLEKDSGAFTLGVI-----TAPVSDWRFY-----D 635  
Db 2503 LW-----FATVKSLLGKGVMLAVSQGRVQTNVNTANEDCIKVAALVNNAFYLENLHFTIE 2558  
Qy 636 SMYTERYMKLTSTNEEGYETSAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALV----- 691  
Db 2559 GKDTHYFIKT--TTPE-S-DLGTALRTLSGRKALENGI-----NVTVSQSTTVVNGRT 2606  
Qy 692 ----DLIMGDVSPEKLSHQWFTSDHGISVHGGGVFLYKQARKLYQEKNRQTVLMHQ 747  
Db 2607 RRFADEVQFGAL--ALHVR-----GMTLDEEKARILEQAROR---ALARA 2648  
Qy 748 WTK-----KOLEE 755  
Db 2649 WAREQQRVRDGE 2661

RESULT 12  
US-11-096-051-10  
; Sequence 10, Application US/11096051  
; Publication No. US20050244868A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: MacLachlan, Timothy K  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Vernet, Corine  
; APPLICANT: Ethenberg, Seth  
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
; FILE REFERENCE: Attorney Docket No. Cura 967  
; CURRENT APPLICATION NUMBER: US/11/096,051  
; CURRENT FILING DATE: 2005-03-30  
; PRIOR APPLICATION NUMBER: 10/038,854  
; PRIOR FILING DATE: 2001-12-31  
; PRIOR APPLICATION NUMBER: 10/455,772  
; PRIOR FILING DATE: 2003-06-04  
; PRIOR APPLICATION NUMBER: 60/557,978  
; PRIOR FILING DATE: 2004-03-30  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO 10  
; LENGTH: 2721  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-096-051-10  
Query Match 3.1%; Score 123.5; DB 7; Length 2721;  
Best Local Similarity 20.0%; Pred. No. 0.57;  
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

Qy 60 NIVT--NRSRTLI PADKPTGKEA-FNYWIHPDLSSVLWASNHTKQYRHSFPADYVQDV 116  
Db 1945 SIITDYNEGLLQTAFLGTSRRVLFKYNRQTRLSEILYDST-----RVSP---TYDETA 1996

Qy 117 ESLSKSVPLMDQEG---DIOYAQWSPVGNITAF-----VRENDLYVMNGTVTTRITDD 166  
Db 1997 GVLKTVNLQSD--GFICTIRYQIGPLIDROIFRSEDGMVNARFDYSIDNS--FRVTSM 2052  
Qy 167 GGPDMFHGVPDWIYEEBILGDRVALWSPDGEV-LAYLSFNE---TCVPYVT----- 214  
Db 2053 QGVINETPLFDIYQFDDISGRKEQF-----GKGVYIYDINOIISTAVMTYTKHFDAGR 2108  
Qy 215 ---VOYTMDOEIAPAYPWELKIRYKPVQSNPTVTLSLNLNIASKVQKAP----- 262  
Db 2109 IKEIQY-----EIFRSLMYWITIQYDNMGR-----VTKREIKIGPFANTTKYAY 2152  
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDVTASNKATVLSD---R 316  
Db 2153 EYDVGOLQTVYLNKIMRMRYNDLNGNLHLNPSNSARLTPRLYDURDRITRUGDVOYR 2212  
Qy 317 DGTGDL-----DNLLSMKYIGIPKPSDKDAYVIDISDHSQWHLFLFPVSGG 364  
Db 2213 LDEGFLRQRGTETFEYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 2256  
Qy 365 EPIPLTKGDWEVTSILSIDQERQLVYVYLSLSTQHHSRHLHSVSVSTFAVTPLVDDTVAAV 424  
Db 2257 -----RVSSKTSILGQHLQF-FYADLTYPTRITHVY--NHSSEITSLSYLDLQHL 2303  
Qy 425 WSASFANSYGYILTYGGPDVPYQELVYTNSTKPLRTITDNKVLKQIKOYALPNITYFE 484  
Db 2304 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQTAYGEI-YFD 2349  
Qy 485 LPLPSGETLNMORLPPGFS-----PDKKYPILFTPYGPGGAQEVTKR--WQ 529  
Db 2350 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTPDI-EIWKRIKGD 2400  
Qy 530 ALNFKAYVASDSE-----LEYVT---WTV-----DNRGTGFKGRKFRSAVTRQLG 571  
Db 2401 PAPENLYMFRNPNASKIHDVKDYITDVNSLWTFGFHLHNAIPGFPVPKF-----D 2452  
Qy 572 LLEAQDIYAAQAANIPWIDADHIGI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGGAQSW 2508  
Db 2453 LTPESYELVKVSSQWDDIPPI-----VLEKDSGAFTLGVI-----TAPVSDWRFY-----D 635  
Qy 600 GWSFGGYLTSK-----FATVKSLLGKGVMLAVSQGRVQTNVNTANEDCIKVAALVNNAFYLENLHFTIE 2564  
Db 2509 LW-----SMYTERYMKLTSTNEEGYETSAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALV----- 691  
Qy 636 SMYTERYMKLTSTNEEGYETSAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALV----- 691  
Db 2565 GKDTHYFIKT--TTPE-S-DLGTALRTLSGRKALENGI-----NVTVSQSTTVVNGRT 2612  
Qy 692 ----DLIMGDVSPEKLSHQWFTSDHGISVHGGGVFLYKQARKLYQEKNRQTVLMHQ 747  
Db 2613 RRFADEVQFGAL--ALHVR-----GMTLDEEKARILEQAROR---ALARA 2654  
Qy 748 WTK-----KOLEE 755  
Db 2655 WAREQQRVRDGE 2667

## RESULT 13

US-11-096-051-8  
; Sequence 8, Application US/11096051  
; Publication No. US20050244868A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: MacLachlan, Timothy K  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Vernet, Corine  
; APPLICANT: Ethenberg, Seth  
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
; FILE REFERENCE: Attorney Docket No. Cura 967  
; CURRENT APPLICATION NUMBER: US/11/096,051  
; CURRENT FILING DATE: 2005-03-30  
; PRIOR APPLICATION NUMBER: 10/038,854

;; PRIOR FILING DATE: 2001-12-31  
;; PRIOR APPLICATION NUMBER: 10/455,772  
;; PRIOR FILING DATE: 2003-06-04  
;; PRIOR APPLICATION NUMBER: 60/557,978  
;; PRIOR FILING DATE: 2004-03-30  
;; NUMBER OF SEQ ID NOS: 38  
;; SOFTWARE: CuraseqList version 0.1  
;; SEQ ID NO 8  
;; LENGTH: 2725  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-11-096-051-8

Query Match 3.1%; Score 123.5; DB 7; Length 2725;  
Best Local Similarity 20.0%; Pred. No. 0.57;  
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY 60 NIWT--NESRTIIPADKIPTGKEA--FNWIHPDLSSVLWASNHTKQYRHSFPFADYVQDV 116  
DB SIITDYNBEGLLLOTAFLGTSRRVLPKYRQRLSEILYDST----RVSF--TYDETA 2000

QY 117 ESKSVPLMPDQEG---DIQYAGMSVGNNTIAP-----VRENDLYVWDNQTIVTRITD 166  
DB GVLKTNLQSD--GFICTIRYRQIGPLIDRQIFRFSDEGWNARFDYSYDNS--FRVTSM 2056

QY 167 GGPDMFHGVPDWIYEEIILGDYALWFSPDGEY-LAYLSFNE---TGVPYTN----- 214  
DB 2057 QGVINETPLIDLYQDDISGKVEQP---GKFGVIYDINOIISTAVNTYTKGFDHGR 2112

QY 215 ---VOYMDNQIAPAYPWEIKIRYPKVQNTPTVTLNLINIASKEVQAP----- 262  
DB 2113 IKEIQY-----EIFRSLMTWIIQYDNMGR-----VTKREIKIGPANTTKYAY 2156

QY 263 ----IDAFESTDLIGEVAWLTDTHTTAAKAFNRVQDQKVAVDATASKATVISD---R 316  
DB 2157 HYVDVQQLQTVYLNKIMRYNDLGNLHLLNPSNSARLTPLRYLDRITRLGDOYR 2216

QY 317 DGTGDLW-----DNLLSMKYIGPIKPSDKDAYIIDSDHSGWAHLVLPVSGG 364  
DB 2217 LDEDFLQRGTETFEYSSKGLLTRY-----SKSGMTVIYRYDGLGR 2260

QY 365 EPIPLTKGDWEVTSILSDQEROLVYLGSTQHHSTERHLYSVSYSTFAVTPLVDDTVAA 424  
DB 2261 ----RVSSKTSLGHLQF--FYADLTPTRIYV--NHSSEITSLYDLOGLH 2307

QY 425 WSASFSGANGYILTYGGDPVPQELTYNTSTKPLRTITDNNAKLEQIKDYALPNITYPE 484  
DB 2308 FAMEISSGDEFYIA-----SDNTGTPLAFFSSNGLMKLQIQVYAYGEI--YFD 2353

QY 485 LPLPSGETLNMORLPPGFS-----PDKKYPILTFTPYGPGQAQVTKR--WQ 529  
DB 2354 -----SNIDPQLVIFGHGGLYDPLTKLIHFERDYLILAGRTTPOI--EIKRIGKD 2404

QY 530 ALNFKAYVADSE-----LEVVT-----WTV-----DNRGTGFKGRKFRSAVTRQLG 571  
DB 2405 PAPFNLYMERNPNPASKIHVDKDIITDVNSWLVTFGFHLHNAIPGPVPKF-----D 2456

QY 572 LLEAEQDIYAAQQAANIPWIDADHIGI-----FGVQQQVARQAKAFLSLGKMAEVQVSRRAAGQSW 599  
DB 2457 LTPESVELKSKQWDDIPPI-----VLEKDSGAFTLGI-----TAPVSDWRFY-----D 635

QY 600 GWSFGGYLTSK-----VLEKDSGAFTLGI-----TAPVSDWRFY-----D 635  
DB 2513 LW-----FATVKSILIGKVMIAVSGRQVTNVLNANEDCIIKVAAVLNNAFYLENLHFTIE 2568

QY 636 SMYTERYMKTLSTNEGYETSAVRKTDGFKNVGGFLIOHGTGDDNVHVFQNSAALV---- 691  
DB 2569 GKDTHYFIKT--TTPSS--DIGTLRLTSGRKALENGI-----NWTVSQSTTVNGRT 2616

QY 692 ----DLIMGDGVSPEKLHSGWFTSDSHGISYHGGGVFLYKQARKLYQEKNRQTVLMHQ 747  
DB 2617 RRPADEVMOFGAL--ALHVRV-----GMTLDEKARIIEQAQR--ALARA 2658

QY 748 WTK-----KLEEE 755  
DB 2659 WAREQORVRDGE 2671

RESULT 14  
US-10-995-561-625  
; Sequence 625, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 625  
; LENGTH: 1259  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-995-561-625

Query Match 3.0%; Score 122.5; DB 6; Length 1259;  
Best Local Similarity 21.3%; Pred. No. 0.23;  
Matches 135; Conservative 75; Mismatches 205; Indels 219; Gaps 34;

QY 181 EBEIILGDYALWFS---SPDGEYLAFLSFNETGPTVTVYQYMDNQE-----IAPAYPW 230  
DB 297 EEWVHADQSSCTFDNLSPGLE-----NVSVYTVK---DDKESVPISDTIIPAVPP 344

QY 231 ELKIRYPKVQNTPTVTL-----LLNIASKEVQAPIDAFESTDLIIGEVAWLTDTHTTVA 287  
DB 345 PDLRFTNIGPDMRTVWAPPSSIDLTLNLFVRSVPKNEE-----DVABLISISDINA 397

QY 288 AKAFNRVQDQKVAVDTA--SNKATVISDRDGTGWLNDLNLMSKYGIPKPSDKDAYI 345  
DB 398 VVLTNLLPCTEYVVSVSVEQHESTPLGRQKT-----GLDSPTGID--FS 442

QY 346 DISHSGWAHLVFP---VSG-----GEP---IP-----LTGQWEVT 377  
DB 443 DITANSFTVH--WIAPRATITGYRIHHPHFGSRPREDRVPHSRNSITLTLNLTGTEYV 501

QY 378 SILSDIOERQVYLLSTQHHSTERHLYSVSYSTFAVTPLVDDTVAA-----YWSASFS 430  
DB 502 SIVALNGREESPLLIGQQ-----STVDVPRDLEVVAAATPSSLISWDAP-A 547

QY 431 ANSGYVILTYG--GPDVPYQELTYNTSTKPLRTITDNNAKLEQIKDYALP----- 478  
DB 548 VTVRYRYITTYGTTGNSPVQE--FTVPGSKSTATISG---LKPQVDYITTVYAVTGRGDS 602

QY 479 -----NTTY--FELPLPSGETLNMWQ-----RLPPGSPDKKYPILTFTPYGPGQAQEV 524  
DB 603 PASSKPEISINRYTEIDKPSQOMQVTDQDNSISVKMLPSSSPVTVGYRVTTPKNGPPTKT 662

QY 525 TKRQALNFKAIVASDESELEYVTVVDNRGTGFKGRKFRSAVT---RQLGLLEAEQDIYA 581  
DB 663 KTAGPDQTEMTIEGLQPTVEYVSVYAQNPSGESQPLVQTAVTNIDRPKGLAFTD----- 717

QY 582 AQAANIPIWDADHIGIWMGSGFYLTSLVLEKDSGAFTLGVITAPVSDWRFRYDSMYTER 641  
DB 718 -----VDVDSIKI--AW-----ESPQG-----QVSR 736

QY 642 YMKLTSTNBEQ-----YETSAVRKTDGFKRNVGGLIOHQ 676  
DB 737 YRVYSSPEDGIELFPADGEBDTABLQRLPGSEYTVSWALHD---DMESQPLI--G 791

QY 677 TGDNDVHFQNSA--ALVDLLMGDGVSPKLSQW 708  
DB 792 T-----OSTAIPAPTDLKFTQ--VTPTSLSAQW 817

```
RESULT 15
US-10-995-561-628
; Sequence 628, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 628
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-628

Query Match      3.0%; Score 122.5; DB 6; Length 1286;
Best Local Similarity 21.3%; Pred. No. 0.23; Mismatches 205; Indels 219; Gaps 34;
Matches 135; Conservative 75;

Qy 181 EEEILGDRYALWF---SPDGEYLAYLSFNETGVPTTYVQYMDNQE-----IAPAYPW 230
Db 297 EEVVAHQSSCTFDNLSFLEY-----NVSYYTVK---DDKESVPISDTIIPAVPP 344

Qy 231 ELKIRYKVSQTNPTVTLUS-----LNIASKEVKQAPIDAFESTDLIIGEVAMLTHTHTVA 287
Db 345 PTDLRFTNIGPDTWRVTWAPPPSIDLTNFLVRYSPVKNEE-----DVAELSI SPSDNA 397

Qy 288 AKAFNRVQDQKVAVDTA--SNKATVISDRDGTGMDLNLMSKYIGPIKPSDKDAYI 345
Db 398 VVLNLLPGTEYVSVSVVYQHESTPLRGQKT-----GLDSPTGID--FS 442

Qy 346 DISHSGWAHLYLFP---VSG-----GEP---IP-----LTKGDM EVT 377
Db 443 DITANSFTVH-WIAPRAITGVRIRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYVV 501

Qy 378 SILSDQERQLVYVYLSQHHSTERHLYSVSSTFAVTLVDDTVAA-----YWSASF 430
Db 502 SIVALNGREESPLLTGQQ-----STVSDVPRDLEVAATPTSLLSMDAP-A 547

Qy 431 ANSGYVILTYG--GPDVPYQELTYTNTSTKPLRTITDNAKVLEQIKDYALP----- 478
Db 548 VTVRYRITYGETGNSPVQE--FTVPGSKSTATISG----LXPGVDYITVYAVTGRGDS 602

Qy 479 -----NITY-FELPLPSGETLNMVQ-----RLPPGFSPPKYPILFTPYGPGAEV 524
Db 603 PASSKPI SINRYTEIDKFSQMQVTDVQNSISVKWLPSSPVTVGYRVTTTPKNGPGPTKT 662

Qy 525 TKRQALNFKAVASDSLELYVTWTVDNRGTGFKGKFRSAVT---ROLGLLEAEDQIYA 581
Db 663 KTAGPDQTEMTIEGLQPTVEYVSVYAQNPSGESQPLVQTAVTNIDRPKGLAFTD----- 717

Qy 582 AQQAANIPIWDADHIGIWCWSFGGYLTSKVLEKDSGAFTLGVITAPVSDWRFPYDSMYTER 641
Db 718 -----VDVDSIKI-AW-----ESPOG-----QVSR 736

Qy 642 YMKTLTSTNEEG-----YETSAVRKTDGFKNVGEGFLIQHG 676
Db 737 YRVTYSSPEDGIHELFPAPDGEEDTAELQGLRPGSEYTVSVVALHD---DMESQPLI--G 791

Qy 677 TGDDNVHFQNSA--ALVDLLMGDGVSPKLSQW 708
Db 792 T-----QSTAIPAPTDLKFTQ-VTPTSLSAQW 817
```